Ukrainian Catholic University

Faculty of Applied Sciences

Business Analytics & Computer Science Programmes

A guide for beginner musicians:
How to become popular in the modern world

Econometrics final project report

Authors:

Anastasiia Kurylets,
Julia Chebotarova,
Viktoriia Markus

Contents:

Introduction......3
Literature review......3
Data description......3
Data analysis:......4
Methods......5
Results......6
Conclusions......6
Appendixes:......9

Introduction
Becoming popular in today's music world demands more than talent — it requires smart strategy and constant adaptation.
The aim of this research is to explore how aspiring musicians can achieve popularity in the modern world by analyzing the characteristics of popular music. Using Spotify data we identify possible practical strategies related to musical style and audience engagement.

Literature review
1. Angrist, J. D., & Pischke, J.-S. (2008). Mostly Harmless Econometrics.
From this source, we used the concept of omitted variable bias, which underlines the importance of including all relevant predictors in regression analysis. The book also emphasizes that ordinary least squares estimators are only unbiased when all Gauss-Markov assumptions hold, which guided our decision to test model assumptions before interpreting results.

2. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An Introduction to Statistical Learning.
This work provided a structured overview of regression modeling and diagnostics. We used it specifically for understanding model evaluation metrics such as R-squared and RMSE; learning about assumption tests; recognizing the importance of proper model selection and interpretation in statistical learning.

3. Wooldridge, J. M. (2015). Introductory Econometrics: A Modern Approach.
This book helped us clarify the correct interpretation of coefficients in linear models and provided practical examples of how multicollinearity and omitted variables can distort inference. It reinforced the need to check variance inflation factors when selecting independent variables.

Data description
For our research, we decided to choose the "30000 Spotify Songs". This dataset contains information about 30 000 songs available on Spotify. The data is obtained through the Spotify API and covers a variety of musical characteristics, allowing us to analyze the factors that influence the popularity of tracks.

Variables:
track_id: Unique identifier for the song.
track_name: Name of the song.

track_artist: Artist who performed the song.

track_album_id: Unique identifier for the album.

track_album_name: Name of the album the song is from.

track_album_release_date: Release date of the album.

playlist_name: Name of the playlist containing the song.

playlist_id: Unique identifier for the playlist.

We do not take it into account variables above, because our main goal is to understand what properties of the song itself make it popular.

Selected variables:
track_popularity: Popularity score of the song (0–100), with higher values indicating greater popularity.

playlist_genre: Genre classification of the playlist.

playlist_subgenre: More specific subgenre of the playlist.

danceability: Score (0.0–1.0) indicating how suitable the song is for dancing.

energy: Score (0.0–1.0) reflecting the track's intensity and activity.

key: Musical key of the track, represented as integers (0 = C, 1 = C♯/D♭, etc.).

loudness: Average loudness of the track in decibels (dB).

mode: Indicates if the track is in a major (1) or minor (0) key.

speechiness: Score (0.0–1.0) indicating the presence of spoken words.

acousticness: Confidence score (0.0–1.0) that the track is acoustic.
instrumentalness: Likelihood (0.0–1.0) that the track has no vocals.

liveness: Probability (0.0–1.0) that the track was recorded live.

valence: Score (0.0–1.0) representing musical positivity or mood.

tempo: Estimated tempo in beats per minute (BPM).

duration_ms: Length of the track in milliseconds.

Data analysis:

Missing values:

On the picture 1 (Appendixes, data analysis) you can see that the dataset contains almost no missing values. This allows you to proceed to data analysis and model building without additional cleaning. The completeness of the data ensures higher accuracy of further statistical conclusions.

Distribution of songs:

The plot 2 (Appendixes, data analysis) shows how the popularity of approximately 30,000 songs on Spotify is distributed on a scale from 0 to 100. Most tracks are popular between 40-65 points, where the second highest bar is visible. This suggests a concentration of moderately popular songs in that range.
It is also evident that relatively few songs achieve very high popularity (above 85), while a broad distribution spans from 20 to 80. In summary, most songs on Spotify maintain average to moderately high popularity, with only a few becoming viral hits but many remaining entirely obscure.

Multicollinearity:

The plot 3 (Appendixes, data analysis) presents the correlation matrix heatmap of all numerical independent variables used in the model. It helps to visually detect multicollinearity, i.e. strong linear relationships between predictors. While most variable pairs show low or moderate correlation, a few noteworthy patterns are observed. For instance, energy and loudness are positively correlated ($r = 0.68$), suggesting that more energetic songs tend to be louder. Another significant relationship is the strong negative correlation between energy and acousticness ($r = –0.54$), which reflects that highly energetic songs are less likely to be acoustic. Additionally, danceability and valence show a mild positive correlation ($r = 0.33$).

To formally detect multicollinearity among the independent variables, we computed the Variance Inflation Factor (VIF) for each predictor. The results are shown in the picture 4 (Appendixes, data analysis). All VIF values are below 3, with the highest being for energy (VIF = 2.66) and loudness (VIF = 2.08), which aligns with their moderately strong correlation in the heatmap (plot 3). The intercept (const) naturally shows a very high VIF, which is expected and not a concern.

Methods

1. Ordinary Least Squares (OLS) Regression

To investigate the factors that influence a song's popularity on Spotify, we used Ordinary Least Squares (OLS) regression as our primary estimation method. The dependent variable in our model is track_popularity, which ranges from 0 to 100 and reflects how well a song performs on the platform. As independent variables, we included 11 musical characteristics such as danceability, energy, loudness, speechiness, valence, and others.
We chose Ordinary Least Squares (OLS) regression because it provides interpretable estimates of how each musical feature affects song popularity. It allows us to formally test hypotheses using t-tests and F-tests, and under standard assumptions, OLS produces efficient and unbiased results. This method is widely used in econometrics, making our analysis both rigorous and accessible.

2. Logit Regression (Binary Logistic Regression)

We used logistic regression to model the probability that a song becomes highly popular, defining a binary variable where 1 indicates a “hit” (popularity > 80) and 0 otherwise. This method is suitable for binary outcomes and allows us to estimate how musical features influence the likelihood of success. The model includes the same predictors as the OLS regression, such as danceability, energy, and valence. Logistic regression was chosen because it provides interpretable results in terms of odds and is well-suited for identifying which characteristics increase the chances of a song becoming a hit.

Results

As a result of the study, we found that:

- For the general dataset, both the OLS model and the Logit model have a low R-squared (7.2%) and Pseudo R-squared (5.4%), i.e. the models are weak, but significant due to other indicators. In general, the popularity of songs on Spotify increases with danceability, loudness, valence and tempo, but decreases with energy, speechiness, instrumentalness, liveliness, and duration, as confirmed by both regressions, although acousticness was significant only in OLS.
- For Pop, OLS ($R^2 = 5.7\%$) and Logit (Pseudo $R^2 = 4.1\%$) show that danceability, energy, speechiness, and acousticness increase popularity, while instrumentalness, valence, and duration reduce it.
- For Rap, OLS ($R^2 = 4.8\%$) and Logit (Pseudo $R^2 = 5.2\%$) confirm that danceability and acousticness increase popularity, whereas speechiness, valence, instrumentalness, and duration decrease it.
- For EDM, OLS ($R^2 = 11.8\%$) and Logit (Pseudo $R^2 = 9.6\%$) show strong influence of acousticness and valence (positive), and instrumentalness, speechiness, tempo, and duration (negative); danceability is not significant in OLS but has a negative effect in Logit.
- For R&B, OLS ($R^2 = 7.1\%$) and Logit (Pseudo $R^2 = 5.6\%$) reveal that energy and acousticness positively affect popularity, while valence_danceability, instrumentalness, liveness, and duration lower it; speechiness and tempo are not significant.
- For Latin, OLS ($R^2 = 6.5\%$) and Logit (Pseudo $R^2 = 5.6\%$) indicate that danceability, energy_loudness, acousticness, speechiness, and tempo raise popularity, while instrumentalness, liveness, and valence reduce it; duration is significant only in OLS.

You can look at the outputs for each model in Appendixes.

Conclusions

We analyzed the impact of musical characteristics on the popularity of songs on Spotify. The main goal was to identify strategies that aspiring musicians can use to gain popularity in today's music industry. By applying econometric models such as OLS and Logit, we identified several important trends.

In general, we found that song popularity increases with higher danceability, loudness, positivity (valence), and tempo, while it decreases with higher energy, speechiness, instrumentalness, liveness, and longer song duration. Although the models had relatively low R-squared values, the results were statistically significant and highlighted key factors influencing success.

When analyzing different genres, we observed specific patterns. In pop music, songs that are danceable, energetic, include spoken elements, and have acoustic qualities tend to be more popular, while those with excessive instrumentalness, high positivity, and long duration tend to be less successful. In rap, danceability and acousticness positively influence popularity, but too much speechiness, positivity, instrumentalness, and longer duration can reduce a song's success. For EDM, acousticness and valence help boost popularity, while too much instrumental focus, high tempo, speechiness, and long duration have a negative effect. In R&B, energy and acousticness support popularity, but tracks that are overly cheerful, dance-heavy, instrumental, or long tend to perform worse. Finally, in Latin music, danceability, energy, loudness, acousticness, speechiness, and a fast tempo are all beneficial, while too much instrumentalness, liveness, and positivity can reduce a song's appeal.

Based on these findings, we suggest that musicians tailor their approach depending on their genre. For pop artists, focusing on catchy, energetic, and moderately acoustic songs is key, avoiding overly long or purely instrumental tracks. Rap musicians should balance rhythm and melody, being careful not to rely too heavily on speech alone and adding acoustic touches to enhance appeal. EDM producers should aim for emotionally uplifting tracks with a clean balance between electronic sounds and vocals, keeping tempo and instrumental sections in moderation. R&B artists are encouraged to create powerful, acoustic-rich songs while avoiding too much cheerfulness or dance emphasis. Latin musicians should prioritize energetic, danceable tracks with strong acoustic and rhythmic elements, while minimizing overuse of instrumental or live performance features.

In conclusion, while talent remains crucial, a thoughtful combination of musical characteristics, adapted to the genre, can significantly enhance an artist's chances of gaining popularity in the modern music scene.

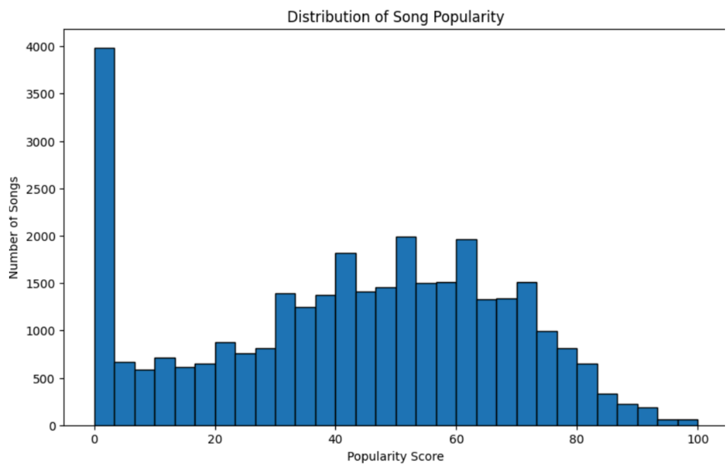
Appendixes:

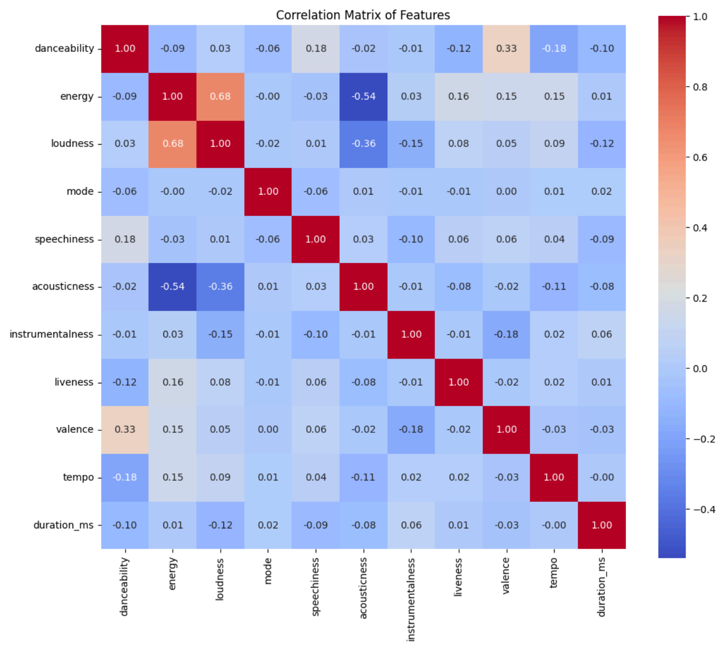
Data Analysis

Picture 1

```
Missing values per column:
track_id                    0
track_name                  5
track_artist                5
track_popularity            0
track_album_id              0
track_album_name            5
track_album_release_date    0
playlist_name               0
playlist_id                 0
playlist_genre              0
playlist_subgenre           0
danceability                0
energy                      0
key                         0
loudness                    0
mode                        0
speechiness                 0
acousticness                0
instrumentalness            0
liveness                    0
valence                     0
tempo                       0
duration_ms                 0
```

Picture 2





Picture 3

	Variable	VIF
0	const	156.463325
1	danceability	1.311575
2	energy	2.661166
3	loudness	2.083386
4	mode	1.007601
5	speechiness	1.069859
6	acousticness	1.464190
7	instrumentalness	1.132005
8	liveness	1.049505
9	valence	1.266500
10	tempo	1.068981
11	duration_ms	1.051185

Picture 4

Models:

1. General Data

OLS Regression Results

Dep. Variable:	track_popularity	R-squared:	0.072
Model:	OLS	Adj. R-squared:	0.072
Method:	Least Squares	F-statistic:	232.1
Date:	Fri, 09 May 2025	Prob (F-statistic):	0.00
Time:	19:52:20	Log-Likelihood:	-1.5102e+05
No. Observations:	32833	AIC:	3.021e+05
Df Residuals:	32821	BIC:	3.022e+05
Df Model:	11		
Covariance Type:	nonrobust		

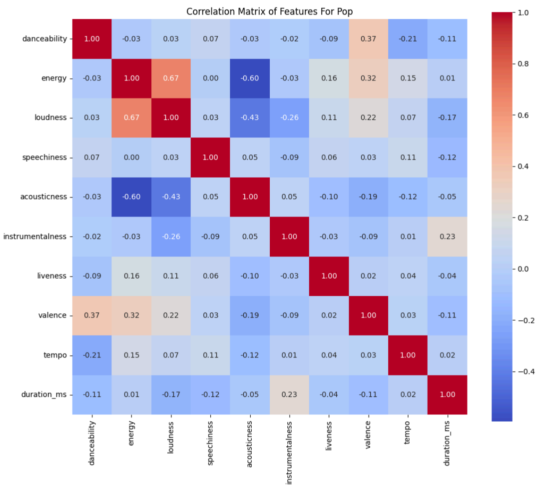
	coef	std err	t	P>|t|	[0.025	0.975]
const	78.0638	1.662	46.982	0.000	74.807	81.321
danceability	5.0361	1.049	4.803	0.000	2.981	7.091
energy	-29.5015	1.198	-24.629	0.000	-31.849	-27.154
loudness	1.5243	0.064	23.758	0.000	1.399	1.650
mode	0.6635	0.269	2.466	0.014	0.136	1.191
speechiness	-7.2310	1.356	-5.332	0.000	-9.889	-4.573
acousticness	3.2319	0.732	4.416	0.000	1.797	4.666
instrumentalness	-11.9744	0.630	-18.998	0.000	-13.210	-10.739
liveness	-4.3131	0.882	-4.891	0.000	-6.042	-2.585
valence	2.7970	0.641	4.362	0.000	1.540	4.054
tempo	0.0212	0.005	4.156	0.000	0.011	0.031
duration_ms	-4.587e-05	2.28e-06	-20.152	0.000	-5.03e-05	-4.14e-05

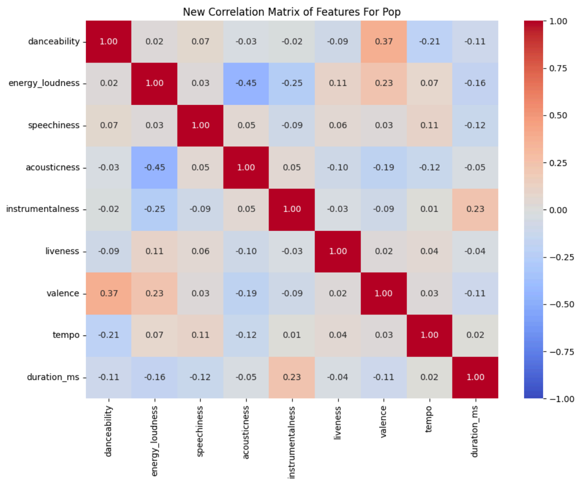
Logit Regression Results

Dep. Variable:	popular	No. Observations:	32833
Model:	Logit	Df Residuals:	32822
Method:	MLE	Df Model:	10
Date:	Fri, 09 May 2025	Pseudo R-squ.:	0.05358
Time:	19:52:21	Log-Likelihood:	-18655.
converged:	True	LL-Null:	-19711.
Covariance Type:	nonrobust	LLR p-value:	0.000

	coef	std err	z	P>|z|	[0.025	0.975]
const	1.9507	0.163	11.996	0.000	1.632	2.269
danceability	0.3084	0.101	3.047	0.002	0.110	0.507
energy	-2.6944	0.118	-22.891	0.000	-2.925	-2.464
loudness	0.1650	0.007	24.591	0.000	0.152	0.178
speechiness	-0.8316	0.129	-6.437	0.000	-1.085	-0.578
acousticness	0.0548	0.069	0.799	0.424	-0.080	0.189
instrumentalness	-1.9380	0.092	-21.053	0.000	-2.118	-1.758
liveness	-0.3088	0.087	-3.534	0.000	-0.480	-0.138
valence	0.5040	0.062	8.119	0.000	0.382	0.626
tempo	0.0018	0.000	3.702	0.000	0.001	0.003
duration_ms	-1.473e-06	2.38e-07	-6.203	0.000	-1.94e-06	-1.01e-06

2. For genre Pop





OLS Regression Results

Dep. Variable:	track_popularity	R-squared:	0.057
Model:	OLS	Adj. R-squared:	0.055
Method:	Least Squares	F-statistic:	36.61
Date:	Fri, 09 May 2025	Prob (F-statistic):	1.34e-63
Time:	20:48:08	Log-Likelihood:	-25414.
No. Observations:	5507	AIC:	5.085e+04
Df Residuals:	5497	BIC:	5.091e+04
Df Model:	9		
Covariance Type:	nonrobust		

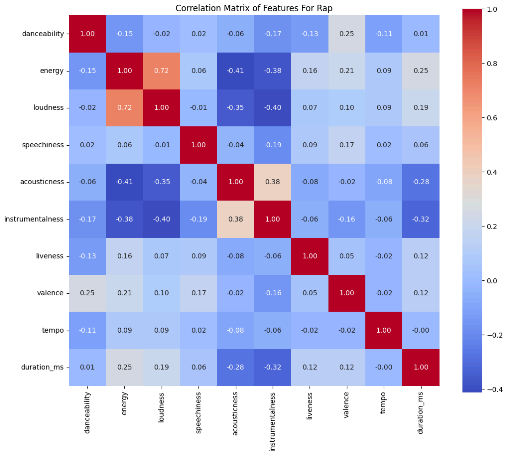
	coef	std err	t	P>|t|	[0.025	0.975]
const	56.6175	3.363	16.834	0.000	50.024	63.211
danceability	17.9317	2.889	6.207	0.000	12.268	23.595
energy_loudness	2.2392	0.286	7.821	0.000	1.678	2.800
speechiness	20.4675	4.974	4.115	0.000	10.717	30.218
acousticness	7.4135	1.727	4.292	0.000	4.027	10.800
instrumentalness	-13.7755	1.901	-7.247	0.000	-17.502	-10.049
liveness	-2.4541	2.460	-0.997	0.319	-7.278	2.369
valence	-4.8624	1.678	-2.898	0.004	-8.151	-1.574
tempo	-0.0255	0.014	-1.833	0.067	-0.053	0.002
duration_ms	-4.613e-05	7.67e-06	-6.010	0.000	-6.12e-05	-3.11e-05

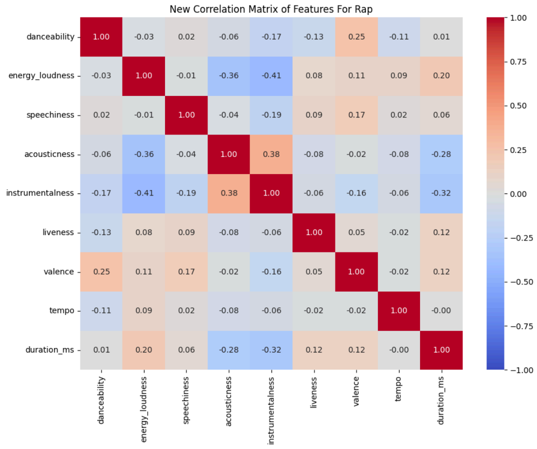
Logit Regression Results

Dep. Variable:	popular	No. Observations:	5507
Model:	Logit	Df Residuals:	5497
Method:	MLE	Df Model:	9
Date:	Fri, 09 May 2025	Pseudo R-squ.:	0.04092
Time:	20:49:55	Log-Likelihood:	-3528.8
converged:	True	LL-Null:	-3679.3
Covariance Type:	nonrobust	LLR p-value:	1.534e-59

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.2117	0.298	-0.710	0.478	-0.796	0.373
danceability	1.3346	0.255	5.240	0.000	0.835	1.834
energy_loudness	0.2329	0.027	8.698	0.000	0.180	0.285
speechiness	1.4367	0.416	3.452	0.001	0.621	2.252
acousticness	0.4586	0.151	3.029	0.002	0.162	0.755
instrumentalness	-1.9704	0.235	-8.385	0.000	-2.431	-1.510
liveness	-0.1490	0.212	-0.703	0.482	-0.564	0.266
valence	-0.3371	0.147	-2.293	0.022	-0.625	-0.049
tempo	-0.0012	0.001	-0.978	0.328	-0.004	0.001
duration_ms	-9.966e-07	7.06e-07	-1.412	0.158	-2.38e-06	3.87e-07

3. For genre Rap

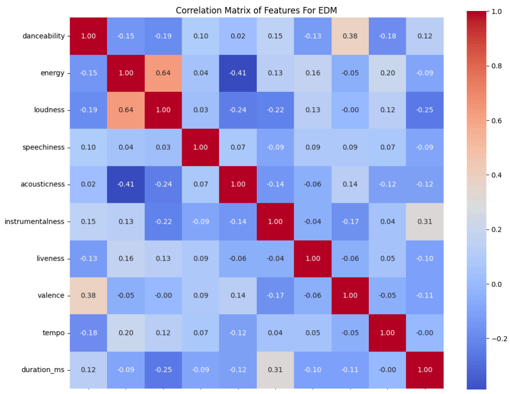


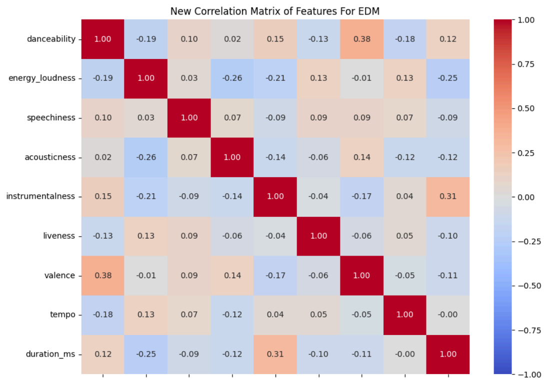



```
                            OLS Regression Results
==============================================================================
Dep. Variable:       track_popularity   R-squared:                       0.048
Model:                            OLS   Adj. R-squared:                  0.047
Method:                 Least Squares   F-statistic:                     32.44
Date:                Fri, 09 May 2025   Prob (F-statistic):           4.03e-56
Time:                        20:52:18   Log-Likelihood:                -26102.
No. Observations:                5746   AIC:                         5.222e+04
Df Residuals:                    5736   BIC:                         5.229e+04
Df Model:                           9
Covariance Type:            nonrobust
====================================================================================
                       coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------------
const               30.6139      2.705     11.317      0.000      25.311      35.917
danceability        27.5948      2.365     11.669      0.000      22.959      32.231
energy_loudness      0.0800      0.217      0.368      0.713      -0.346       0.506
speechiness        -12.0644      2.361     -5.109      0.000     -16.693      -7.435
acousticness         5.6258      1.561      3.605      0.000       2.566       8.685
instrumentalness    -3.0453      1.590     -1.916      0.055      -6.162       0.071
liveness            -0.7385      2.062     -0.358      0.720      -4.781       3.304
valence             -4.3631      1.425     -3.062      0.002      -7.157      -1.569
tempo                0.0485      0.010      5.043      0.000       0.030       0.067
duration_ms      -4.175e-05   5.41e-06     -7.722      0.000   -5.23e-05   -3.11e-05
==============================================================================
```

```
                           Logit Regression Results
==============================================================================
Dep. Variable:                popular   No. Observations:                 5746
Model:                          Logit   Df Residuals:                     5736
Method:                           MLE   Df Model:                            9
Date:                Fri, 09 May 2025   Pseudo R-squ.:                 0.05225
Time:                        20:52:25   Log-Likelihood:                -3138.5
converged:                       True   LL-Null:                       -3311.5
Covariance Type:            nonrobust   LLR p-value:                 4.229e-69
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
const               -3.1275      0.302    -10.354      0.000      -3.719      -2.535
danceability         3.5907      0.270     13.314      0.000       3.062       4.119
energy_loudness      0.1329      0.024      5.541      0.000       0.086       0.180
speechiness         -0.8886      0.253     -3.511      0.000      -1.385      -0.393
acousticness         0.5004      0.165      3.037      0.002       0.177       0.823
instrumentalness    -1.2776      0.210     -6.078      0.000      -1.690      -0.866
liveness             0.1413      0.217      0.653      0.514      -0.283       0.566
valence             -0.6607      0.147     -4.498      0.000      -0.949      -0.373
tempo                0.0044      0.001      4.262      0.000       0.002       0.006
duration_ms      -9.315e-07   5.62e-07     -1.656      0.098   -2.03e-06    1.71e-07
====================================================================================
```

4. For genre EDM (electronic dance music)

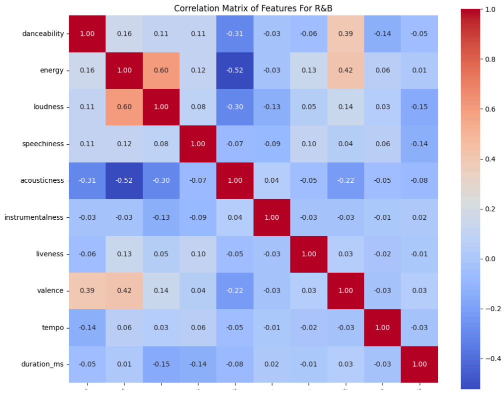


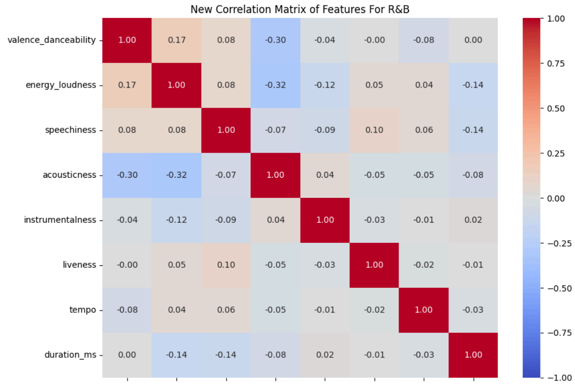



```
                           Logit Regression Results
==============================================================================
Dep. Variable:                popular   No. Observations:                 6043
Model:                          Logit   Df Residuals:                     6033
Method:                           MLE   Df Model:                            9
Date:                Fri, 09 May 2025   Pseudo R-squ.:                 0.09557
Time:                        20:56:15   Log-Likelihood:                -2317.4
converged:                       True   LL-Null:                       -2562.3
Covariance Type:            nonrobust   LLR p-value:                8.986e-100
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.5987      0.424      1.411      0.158      -0.233       1.431
danceability        -0.7370      0.352     -2.096      0.036      -1.426      -0.048
energy_loudness      0.0127      0.036      0.357      0.721      -0.057       0.082
speechiness         -1.9640      0.551     -3.563      0.000      -3.044      -0.884
acousticness         1.5029      0.234      6.425      0.000       1.044       1.961
instrumentalness    -2.2610      0.201    -11.270      0.000      -2.654      -1.868
liveness            -0.4285      0.235     -1.823      0.068      -0.889       0.032
valence              1.1618      0.191      6.068      0.000       0.787       1.537
tempo               -0.0067      0.002     -2.958      0.003      -0.011      -0.002
duration_ms      -5.041e-06   8.26e-07     -6.107      0.000   -6.66e-06   -3.42e-06
```


```
                           Logit Regression Results
==============================================================================
Dep. Variable:                popular   No. Observations:                 6043
Model:                          Logit   Df Residuals:                     6033
Method:                           MLE   Df Model:                            9
Date:                Fri, 09 May 2025   Pseudo R-squ.:                 0.09557
Time:                        20:56:15   Log-Likelihood:                -2317.4
converged:                       True   LL-Null:                       -2562.3
Covariance Type:            nonrobust   LLR p-value:                8.986e-100
====================================================================================
                       coef    std err          z      P>|z|      [0.025      0.975]
------------------------------------------------------------------------------------
const                0.5987      0.424      1.411      0.158      -0.233       1.431
danceability        -0.7370      0.352     -2.096      0.036      -1.426      -0.048
energy_loudness      0.0127      0.036      0.357      0.721      -0.057       0.082
speechiness         -1.9640      0.551     -3.563      0.000      -3.044      -0.884
acousticness         1.5029      0.234      6.425      0.000       1.044       1.961
instrumentalness    -2.2610      0.201    -11.270      0.000      -2.654      -1.868
liveness            -0.4285      0.235     -1.823      0.068      -0.889       0.032
valence              1.1618      0.191      6.068      0.000       0.787       1.537
tempo               -0.0067      0.002     -2.958      0.003      -0.011      -0.002
duration_ms      -5.041e-06   8.26e-07     -6.107      0.000   -6.66e-06   -3.42e-06
```

5. For genre R&B (rhythm and blues)





OLS Regression Results

Dep. Variable:	track_popularity	R-squared:	0.071
Model:	OLS	Adj. R-squared:	0.069
Method:	Least Squares	F-statistic:	51.62
Date:	Fri, 09 May 2025	Prob (F-statistic):	4.47e-81
Time:	21:09:07	Log-Likelihood:	-25179.
No. Observations:	5431	AIC:	5.038e+04
Df Residuals:	5422	BIC:	5.044e+04
Df Model:	8		
Covariance Type:	nonrobust		

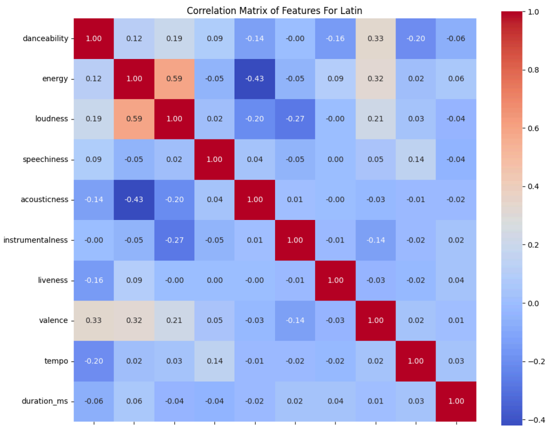
	coef	std err	t	P>|t|	[0.025	0.975]
const	78.2374	2.833	27.618	0.000	72.684	83.791
valence_danceability	-18.2441	2.337	-7.805	0.000	-22.826	-13.662
energy_loudness	2.4001	0.245	9.798	0.000	1.920	2.880
speechiness	-4.0392	3.243	-1.246	0.213	-10.396	2.318
acousticness	5.2595	1.469	3.580	0.000	2.379	8.140
instrumentalness	-10.1944	2.815	-3.621	0.000	-15.713	-4.676
liveness	-11.2318	2.435	-4.612	0.000	-16.006	-6.457
tempo	0.0164	0.012	1.371	0.170	-0.007	0.040
duration_ms	-7.504e-05	6.03e-06	-12.435	0.000	-8.69e-05	-6.32e-05

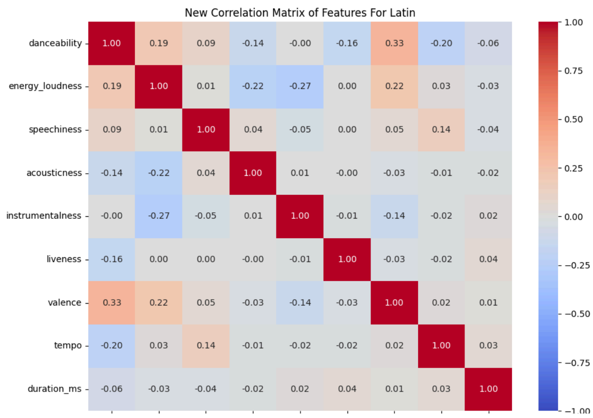
Logit Regression Results

Dep. Variable:	popular	No. Observations:	5431
Model:	Logit	Df Residuals:	5422
Method:	MLE	Df Model:	8
Date:	Fri, 09 May 2025	Pseudo R-squ.:	0.05620
Time:	21:11:48	Log-Likelihood:	-3058.1
converged:	True	LL-Null:	-3240.3
Covariance Type:	nonrobust	LLR p-value:	8.324e-74

	coef	std err	z	P>|z|	[0.025	0.975]
const	2.1677	0.268	8.093	0.000	1.643	2.693
valence_danceability	-1.2203	0.216	-5.654	0.000	-1.643	-0.797
energy_loudness	0.2357	0.025	9.602	0.000	0.188	0.284
speechiness	-0.0595	0.292	-0.204	0.839	-0.632	0.513
acousticness	0.3873	0.134	2.892	0.004	0.125	0.650
instrumentalness	-1.5354	0.346	-4.439	0.000	-2.213	-0.857
liveness	-0.8619	0.240	-3.589	0.000	-1.333	-0.391
tempo	0.0018	0.001	1.628	0.103	-0.000	0.004
duration_ms	-7.113e-06	6.13e-07	-11.597	0.000	-8.31e-06	-5.91e-06

6. For genre Latin





OLS Regression Results

Dep. Variable:	track_popularity	R-squared:	0.065
Model:	OLS	Adj. R-squared:	0.064
Method:	Least Squares	F-statistic:	40.02
Date:	Fri, 09 May 2025	Prob (F-statistic):	1.39e-69
Time:	21:50:02	Log-Likelihood:	-23820.
No. Observations:	5155	AIC:	4.766e+04
Df Residuals:	5145	BIC:	4.773e+04
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	54.6221	3.630	15.048	0.000	47.506	61.738
danceability	9.9475	3.374	2.949	0.003	3.334	16.562
energy_loudness	2.9313	0.254	11.550	0.000	2.434	3.429
speechiness	9.1045	3.992	2.280	0.023	1.278	16.931
acousticness	15.6878	1.661	9.446	0.000	12.432	18.944
instrumentalness	-15.4266	2.128	-7.248	0.000	-19.599	-11.254
liveness	-6.8833	2.309	-2.982	0.003	-11.409	-2.358
valence	-9.2426	1.677	-5.510	0.000	-12.531	-5.954
tempo	0.0279	0.012	2.268	0.023	0.004	0.052
duration_ms	-3.03e-05	7.12e-06	-4.259	0.000	-4.43e-05	-1.64e-05

Logit Regression Results

Dep. Variable:	popular	No. Observations:	5155
Model:	Logit	Df Residuals:	5145
Method:	MLE	Df Model:	9
Date:	Fri, 09 May 2025	Pseudo R-squ.:	0.05648
Time:	21:50:42	Log-Likelihood:	-3172.8
converged:	True	LL-Null:	-3362.7
Covariance Type:	nonrobust	LLR p-value:	2.717e-76

	coef	std err	z	P>|z|	[0.025	0.975]
const	-0.4163	0.326	-1.275	0.202	-1.056	0.223
danceability	0.9502	0.306	3.110	0.002	0.351	1.549
energy_loudness	0.3786	0.028	13.490	0.000	0.324	0.434
speechiness	0.7420	0.348	2.130	0.033	0.059	1.425
acousticness	0.7935	0.152	5.234	0.000	0.496	1.091
instrumentalness	-2.1378	0.319	-6.702	0.000	-2.763	-1.513
liveness	-0.6299	0.211	-2.985	0.003	-1.043	-0.216
valence	-0.4143	0.150	-2.761	0.006	-0.708	-0.120
tempo	0.0023	0.001	2.181	0.029	0.000	0.004
duration_ms	2.771e-07	6.65e-07	0.417	0.677	-1.03e-06	1.58e-06