Uncovering the Beats that move us

Exploring various factors behind the popularity of Hip-Hop Hits

Adrien Picouet   Youness El Khadiri

Ecole Nationale de l'Aviation Civile

Empirical Project
April 24, 2023

Presentation Overview

Context

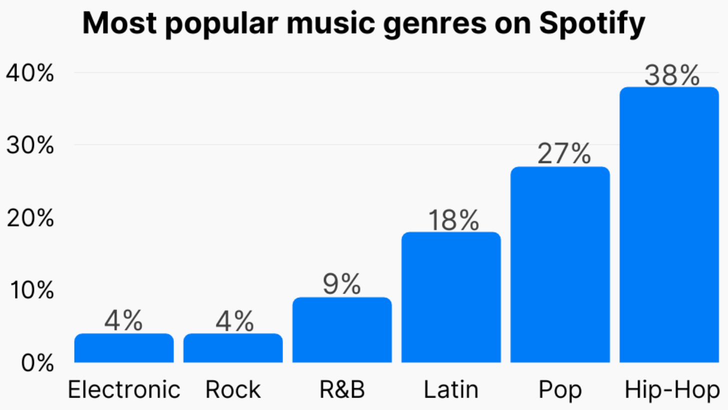


Figure: Spotify US 2022 data

Our Database

Spotify API

Database comes from an API of the leading musical streaming service, Spotify and was gathered via Kaggle.

Sample size

After filtering of the data the sample size is N=547.

Popularity Score : A black box

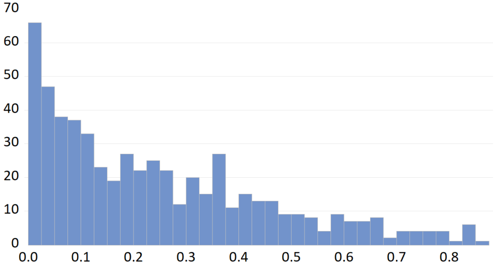
The popularity score (0-100) is calculated by Spotify's algorithm.

Genre Classification

The genre classification may not be perfect due to the subjective nature of music genres.

Variable	Details
ACOUSTICNESS	Acoustic songs use more ”real” instruments
DANCEABILITY	Describes how suitable is a track for dancing
DURATION	Duration of the track in milliseconds
ENERGY	Perceptual measure of intensity and activity
EXPLICIT_TRUE	1 if the track contains explicit content
INSTRUMENTALNESS	Detects presence of vocals
KEY	Key of the track using standard notation
LOUDNESS	Measured acoustic intensity in dB
MODE	1 if the track is in Major mode
SPEECHINESS	Detects presence of spoken words
TEMPO	Tempo of the track in BPM
TIME_SIGNATURE	1x4, 2x4, 3x4 and 4x4

Acousticness



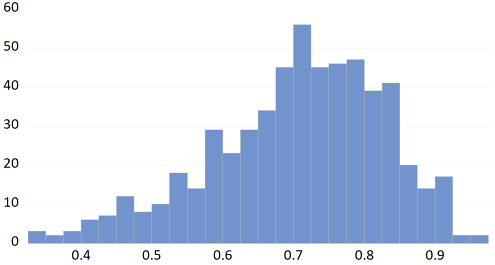
Series: ACOUSTICNESS
Sample 1 573
Observations 572

Mean	0.246272
Median	0.194500
Maximum	0.857000
Minimum	0.000282
Std. Dev.	0.210939
Skewness	0.895341
Kurtosis	3.007146
Jarque-Bera	76.42386
Probability	0.000000

Positive Impact

The use of some real instruments (piano, guitar, drums, flute) in popular tracks is very common.

Danceability

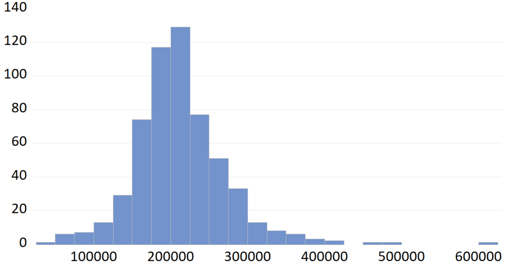


Series: DANCEABILITY	
Sample 1 573	
Observations 572	
Mean	0.705565
Median	0.718500
Maximum	0.955000
Minimum	0.329000
Std. Dev.	0.122199
Skewness	-0.578639
Kurtosis	3.034832
Jarque-Bera	31.94867
Probability	0.000000

Positive Impact

Popular music tends to be joyful and rhythmic.

Duration

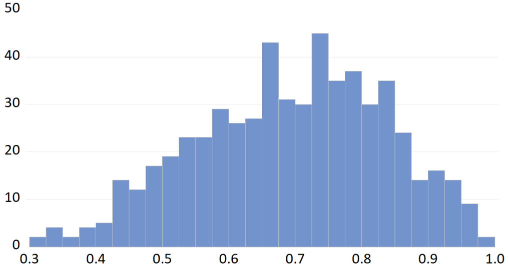


Series: DURATION_MS	
Sample 1 573	
Observations 572	
Mean	212743.3
Median	207746.5
Maximum	620320.0
Minimum	45055.00
Std. Dev.	59752.27
Skewness	1.111598
Kurtosis	8.201757
Jarque-Bera	762.6875
Probability	0.000000

Negative Impact

Longer songs are expected to be less popular.

Energy



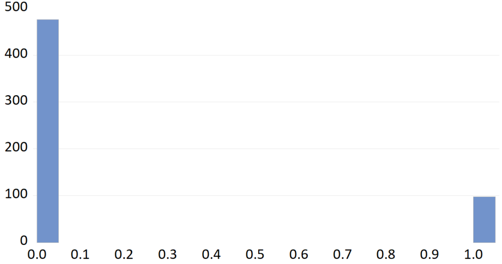
Series: ENERGY
Sample 1 573
Observations 572

Mean	0.693684
Median	0.703500
Maximum	0.994000
Minimum	0.304000
Std. Dev.	0.142465
Skewness	-0.253713
Kurtosis	2.442781
Jarque-Bera	13.53670
Probability	0.001150

Negative Impact

Energy is associated with speed and loudness.

Explicit



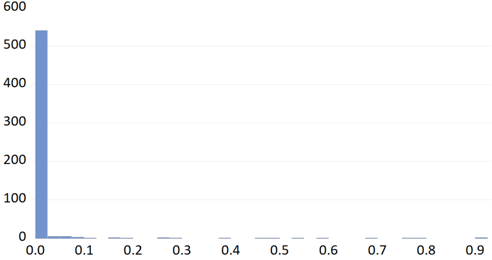
Series: EXPLICIT_TRUE
Sample 1 573
Observations 573

Mean	0.169284
Median	0.000000
Maximum	1.000000
Minimum	0.000000
Std. Dev.	0.375331
Skewness	1.763802
Kurtosis	4.110998
Jarque-Bera	326.5696
Probability	0.000000

Negative Impact

Explicit songs are usually less referenced in playlists, events...

Instrumentalness



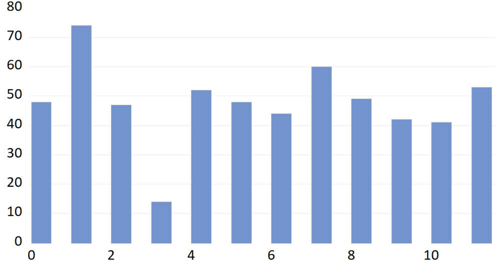
Series: INSTRUMENTALNESS
Sample 1 573
Observations 572

Mean	0.015622
Median	0.000000
Maximum	0.912000
Minimum	0.000000
Std. Dev.	0.091690
Skewness	7.404431
Kurtosis	61.15884
Jarque-Bera	85841.78
Probability	0.000000

Negative Impact

Very few Hip-Hop tracks are considered instrumental.

Key



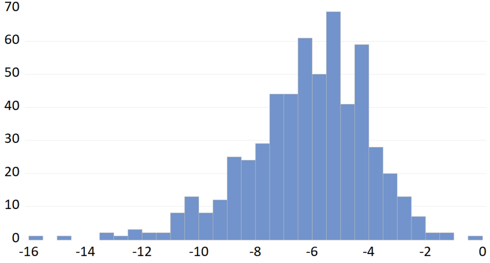
Series: KEY
Sample 1 573
Observations 572

Mean	5.428322
Median	6.000000
Maximum	11.00000
Minimum	0.000000
Std. Dev.	3.536168
Skewness	-0.011126
Kurtosis	1.746583
Jarque-Bera	37.45528
Probability	0.000000

Unknown Impact

It is possible that some keys are more suited for popular tracks.

Loudness



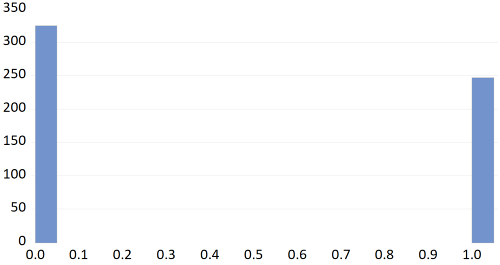
Series: LOUDNESS
Sample 1 573
Observations 572

Mean	-6.158802
Median	-5.930500
Maximum	-0.079000
Minimum	-15.91800
Std. Dev.	2.142324
Skewness	-0.730806
Kurtosis	4.224827
Jarque-Bera	86.67011
Probability	0.000000

Negative Impact

Very loud tracks are not expected to be popular in this genre.

Mode



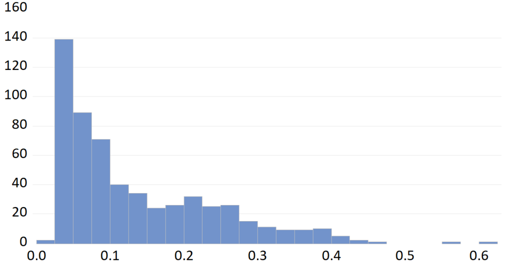
Series: MODE
Sample 1 573
Observations 572

Mean	0.431818
Median	0.000000
Maximum	1.000000
Minimum	0.000000
Std. Dev.	0.495763
Skewness	0.275299
Kurtosis	1.075789
Jarque-Bera	95.47023
Probability	0.000000

Positive Impact

Popular songs are often associated with good mood.

Speechiness

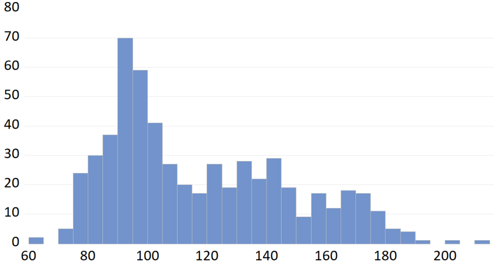


Series: SPEECHINESS	
Sample 1 573	
Observations 572	
Mean	0.133880
Median	0.091200
Maximum	0.621000
Minimum	0.024100
Std. Dev.	0.104302
Skewness	1.198488
Kurtosis	4.051685
Jarque-Bera	163.2949
Probability	0.000000

Unknown impact

Most songs are in the 0-0.1 range

Tempo



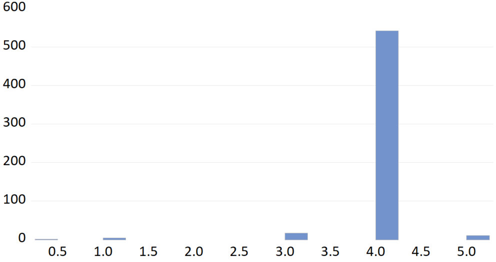
Series: TEMPO
Sample 1 573
Observations 572

Mean	117.4909
Median	109.0090
Maximum	214.0160
Minimum	62.99500
Std. Dev.	30.12967
Skewness	0.621136
Kurtosis	2.378311
Jarque-Bera	45.99209
Probability	0.000000

Unknown impact

Tempo is often used as an indicator of subgenre.

Time Signature



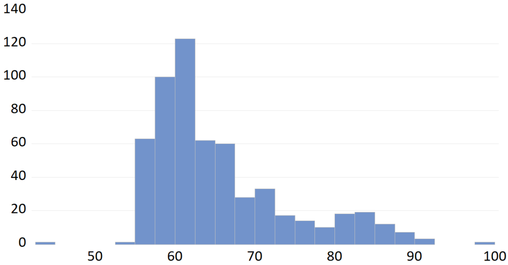
Series: TIME_SIGNATURE
Sample 1 573
Observations 573

Mean	3.962335
Median	4.000000
Maximum	5.000000
Minimum	0.418000
Std. Dev.	0.359725
Skewness	-5.697349
Kurtosis	51.36959
Jarque-Bera	58958.28
Probability	0.000000

Positive impact

Vast majority of tracks with 4/4 time signature.

Popularity



Series: POPULARITY	
Sample 1 573	
Observations 572	
Mean	65.27797
Median	62.00000
Maximum	99.00000
Minimum	43.00000
Std. Dev.	8.547877
Skewness	1.223564
Kurtosis	3.858242
Jarque-Bera	160.2795
Probability	0.000000

Dependant Variable

Low popularity tracks and very high popularity tracks are no very well represented.

Dependent Variable: POPULARITY
Method: Least Squares
Date: 04/24/23 Time: 09:06
Sample: 1 573
Included observations: 572

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	74.03503	4.611915	16.05299	0.0000
ACOUSTICNESS	-5.216179	1.526956	-3.416064	0.0007
DANCEABILITY	4.181314	2.831792	1.476561	0.1404
DURATION_MS	-1.62E-05	5.29E-06	-3.063406	0.0023
ENERGY	-9.178202	2.340842	-3.920897	0.0001
EXPLICIT_TRUE	11.79353	0.828824	14.22923	0.0000
INSTRUMENTALNESS	-1.423983	3.365196	-0.423150	0.6723
LIVENESS	-1.050668	1.966049	-0.534406	0.5933
MODE	-0.122710	0.631158	-0.194420	0.8459
SPEECHINESS	-10.79439	2.992538	-3.607102	0.0003
TEMPO	0.014412	0.010211	1.411359	0.1587
TIME_SIGNATURE	-0.507648	0.923476	-0.549715	0.5827
KEY	-0.012445	0.087528	-0.142181	0.8870
VALENCE	-0.904088	1.609986	-0.561550	0.5746

R-squared	0.344363	Mean dependent var	65.27797
Adjusted R-squared	0.329089	S.D. dependent var	8.547877
S.E. of regression	7.001496	Akaike info criterion	6.754296
Sum squared resid	27353.69	Schwarz criterion	6.860743
Log likelihood	-1917.729	Hannan-Quinn criter.	6.795822
F-statistic	22.54469	Durbin-Watson stat	1.321160
Prob(F-statistic)	0.000000		

Figure: First Regression Results

First Regression
Result Analysis

- Popularity as the dependent variable
- Jointly significant model:
  - F-Statistic [21.49263; 0.00000]
- Model captures more than a third of the variability of the popularity :
  - R^2=0.34
  - Ajusted R^2 =0.33

First Regression

Variable Analysis

- **Significant at 99% level:**
  - Speechiness > Negative impact (importance of lyrics)
  - Explicit_True > Increases popularity by 11,7
  - Energy > Negative impact (hip-hop is not associated with energy)
  - Duration > Negative impact
  - Acousticness > Negative impact (electronic sounds)
- **Significant at 90% level:**
  - Danceability > Positive impact (song played in the club)
  - Key=1 >(C#)
- **Insignificant:**
  - Tempo > Still a good way to differentiate sub-genres
  - Mode > Popular rap songs are not always "joyful"
  - Liveness > the dataset contains few live played songs
  - Instrumentalness > Hip-hop songs are not instrumental
  - Time Signature > Most songs are 4x4

First Regression

Fitted popularity analysis

- Comparison with a perfect regression:
  - Global understimation
  - Large understimation for songs with a popularity > 80

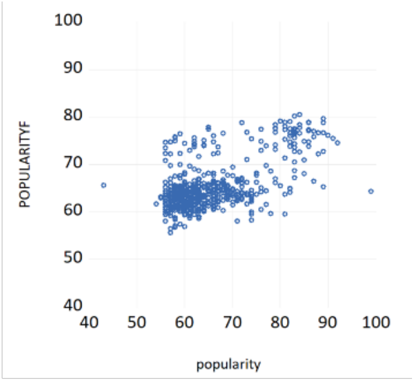


Figure: Fitted popularity plot

Dependent Variable: POPULARITY
Method: Least Squares
Date: 04/24/23 Time: 08:53
Sample: 1 573
Included observations: 572

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	76.25705	6.676490	11.42173	0.0000
ACOUSTICNESS	-5.434199	1.515615	-3.585473	0.0004
DANCEABILITY	4.649815	2.702755	1.720398	0.0859
DURATION_MS	1.14E-05	1.76E-05	0.645441	0.5189
DURATION_MS^2	-5.35E-11	3.42E-11	-1.561566	0.1190
ENERGY	-29.68701	16.86847	-1.759911	0.0790
ENERGY^2	14.65888	12.33472	1.188425	0.2352
EXPLICIT_TRUE	11.98951	0.812940	14.74833	0.0000
SPEECHINESS	-10.80697	2.943995	-3.670852	0.0003
TEMPO	0.013059	0.009980	1.308447	0.1913
TIME_SIGNATURE=4	-1.913584	1.389184	-1.377488	0.1689
KEY=1	1.556367	0.884939	1.758728	0.0792

R-squared	0.353117	Mean dependent var	65.27797
Adjusted R-squared	0.340411	S.D. dependent var	8.547877
S.E. of regression	6.942167	Akaike info criterion	6.733861
Sum squared resid	26988.47	Schwarz criterion	6.825102
Log likelihood	-1913.884	Hannan-Quinn criter.	6.769455
F-statistic	27.79002	Durbin-Watson stat	1.324774
Prob(F-statistic)	0.000000		

Figure: Second Regression Results

Second Model

First regression Improvement

- KEY=1, TIME SIGNATURE=4
- Taking into account factors beyond acoustics
- Studying the impact of popularity by adjusting the tempo variable:
  - Tempo $<$ 100BPM = Rap
  - Tempo $>$ 100BPM = Trap
- Studying the impact of the artist's popularity using the Spotify API:
  - New variable: artist_popularity
- Studying the impact of featuring:
  - New variable: Featuring_true
  - New variable: Ethnicity

Explanatory Variables
Social Factors

Variable	Details
FEATURING_TRUE	1 if the production has a featuring artist
ETHNICITY	1 if the artist is from a minority group
ARTIST_POPULARITY	A score from 0-100 determined from Spotify

Conclusion

Musical Factors

Our results suggest that it is possible to partially explain the popularity of a Hip Hop song by looking at musical parameters from Spotify.

Social Factors

Initial results show that the model can improve by taking in account the popularity of the artist (featuring and ethnicity appear to be insignificant).

Database problem

The database contains a large amount of Indian Hip-Hop, it would be interesting to see how this model performs with a different database.

The End

Questions? Comments?