Data Driven Marketing - Valence Analysis

Baby Goats

```
knitr::opts_chunk$set(warning = FALSE, message = FALSE, echo = FALSE)
```

Background

Music plays an important part in the daily routine of millions of people. Americans in particular are said to listens to an average of 4.5 hours of music a day according to the global measurement company Neilson. Music is essentially something that stimulates the auditory nerves (hearing sense). No two person perceives music in exactly the same way. Very often, the music we listen to influences our mood. Sometimes people turn to music for: Enjoyment, Special emotion, Structural parts of the song which are preferred (chorus, beat, etc.), Unique sensations like ASMR (head orgasm), piloerections (goosebumps), euphoria, deep trance, etc., Social/Interpersonal bonding.

In that sense we realize there are different emotions associated with music. Our analysis revolves around understanding what parameters affects the way a song provokes our emotions. A specific song feature that quantifies the emotional value of a song is called valence. This will be the main topic of our analysis.

Methodology

The plan is to analyze the relationship between valence and the other characteristic of a song, especially for a "positive/happy song". To that effect, we set a cutoff for valence,a value greater than .66 will represent a positive song, and one lower than .33 a sad song. In other terms, one could divide music into happy, less happy (Sad) and neutral - the neutral will be excluded from this analysis // might need to mention why //.

This would allow us to create a binary variable "Positive" (1s and 0s) for a logistic regression. One thing to keep in mind is that the impact of different song features could vary across. So, for the purpose of a complete and accurate analysis, we will be building a model for each of the genres present in our data set.

The data set we will be working with comes from Spotify and has the pop, r&b, edm, latin, rap and rock genres.

Usefulness of Analysis

This model could serve different purpose depending on the nature of the client requesting its access. It would benefit an existing artist trying to improve or sway his music toward a positive edge, just as much as it would benefit a new artist trying to understand what goes into making a happy song in the genre of his choosing.

This would also help producers in reaching out to artists with a strong understanding of what creates a great song.

Data Exploration

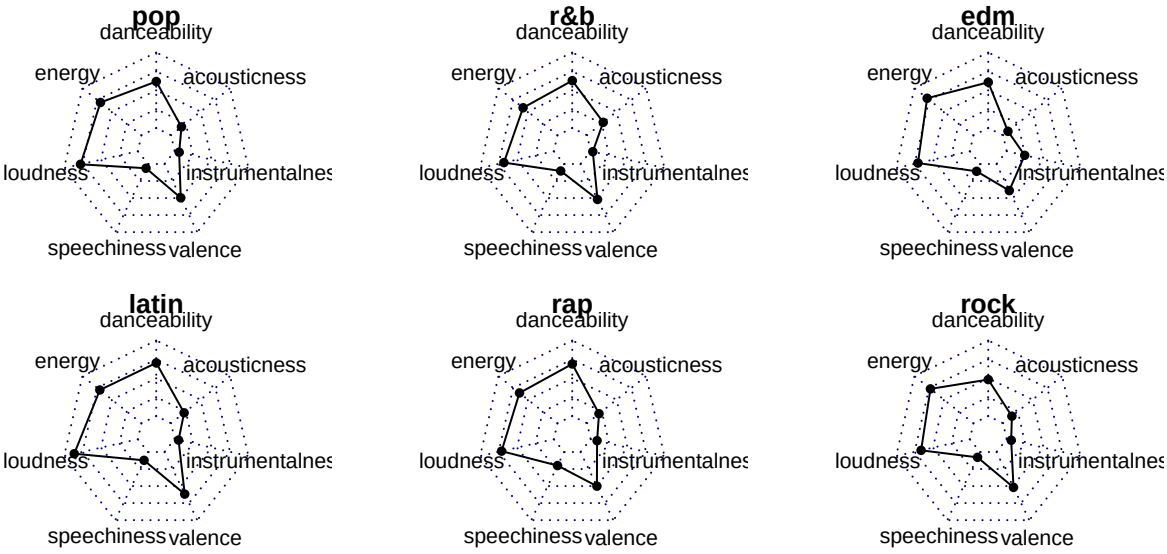
The data set provided contains about 32000 songs with 24 different descriptive characteristics each. The main ones that will be used for this analysis are:

- track_popularity : Measure the popularity from 0 to 100 based on play number of the track
- acousticness : Measure of how acoustic the track is and ranges from 0.0 to 1.0
- danceability : Describes how suitable a track is for dancing. Values range from 0.0 being least danceable and 1.0 being most danceable.
- duration_ms : The duration of the track in milliseconds(ms) which has been converted to minutes using transformation
- energy : Measure from 0.0 to 1.0 and represents a perceptual measure of intensity and activity i.e. the enery of the song.
- instrumentalness : Measure whether a track contains vocals. Sounds are treated as instrumental in this context. Values ranges from 0.0 to 1.0

- speechiness - Detects the presence of spoken words in a track.Values > 0.6 might be a podcast or talk show, where 0.3 to 0.6 is the normal range for songs and if its less than 0.3 its mostly music
- valence - Measure from 0.0 to 1.0 describing the musical positiveness conveyed by a track. Tracks with high valence sound more positive , while tracks with low valence sound more negative.
- key : Estimated overall key of the track. If key is not detected, the value is -1.
- liveness - Detects the presence of an audience in the recording. Higher liveness values represent an increased probability that the track was performed live. A value above 0.8 provides strong likelihood that the track is live.
- loudness - overall loudness of a track in decibels (dB).Values typical range between -60 and 0 dB.
- mode - Mode indicates the modality (major or minor) of a track. Major is represented by 1 and minor is represented by 0.
- tempo Overall estimated tempo of a track in beats per minute (BPM).

Not all of these characteristics are useful in understanding or predicting valence. By example, popularity, especially as designed by Spotify is ***. It would be an important factor to consider in terms fo endogeneity, but would be useless for the predictions we are making.

The radar charts below allow us to compare the different genres on the various features that wil lmake our model.



The models

A correlation plot was used to assess the connections between our dependent variable (Positive set based off the valence levels cutoff) and the other features. To avoid making assumptions based off of correlation alone, we ran one logistic model so as to determine which variables were significant to our analysis and which weren't.

A logistic regression seemed most appropriate as this type of analysis can help predict the likelihood of an event happening or in our case, a song with high valence being produced. It not only provides a measure of how appropriate a predictor(coefficient size) is, but also its direction of association (positive or negative). This is particularly helpful in understanding how shifting the levels of the different features might push a song in one direction or the other (happy or sad).

After removing each insignificant variables for the genres in which they were found as such, we split the data for each group into a training set and a testing set. A second logistic model was then built on the training set and validated on the testing set.

The table below shows a summary of the respective models for each genre.

At a glance, it appears that increasing danceability tends to sway the predictors in a positive direction whereas variables like loudness push it in the negative direction.

Impact

Table 1: Summary of Models per Genres

Features	Pop	RnB	Rap	Rock	Latin	EDM
Intercept	-10.47899882	-12.17745	-11.08675875	-14.06903606	-10.4596	-13.087326
danceability	9.451699067	8.12939	6.859492604	12.61547	8.7169	12.884779
energy	6.86164266	8.95258	5.664478346	8.323076141	6.959	2.292205
key	-	-	-	-	-	-
loudness	-0.08386884	-0.19963	-0.112847895	-0.322133658	-	-
mode1	-	-	-	0.375378468	-	-
speechiness	-	-	3.720055269	-8.133885676	2.3361	-
acousticness	-	1.42444	1.409829896	-	2.4362	1.952887
instrumentalness	-1.389634864	-	-	-1.8395643	-2.0285	-1.523018
tempo	0.005612948	-	-	0.010749491	-	0.012733
duration_ms	-0.000006796	-	0.000004403	-0.000007277	-	-
Model Accuracy	0.7966963	0.8305322	0.7174678	0.8473837	0.8383978	0.7802198

Note:
'-' shows variables insignificant at alpha = 0.05

Averages for Valence in Pop Model

```
##
## 0.488673
```

Danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0985  0.5610  0.6570  0.6411  0.7360  0.9650
```

Table 2: Changing the Danceability

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.2	0.63	0.92

energy

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0436  0.5760  0.7240  0.6885  0.8300  0.9980
```

Table 3: Changing the Energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.21	0.51	0.8

loudness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -26.279  -8.035  -6.043  -6.653  -4.663  -0.700
```

Table 4: Changing the loudness

	-9	-6.5	-4
Predicted Probability of Being Positive	0.54	0.49	0.43

instrumentalness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
```

```
## 0.0000000 0.0000000 0.0000311 0.0744338 0.0061300 0.9660000
```

Table 5: Changing the instrumentalness

	0	0.02	0.05	0.1	0.5	1
Predicted Probability of Being Positive	0.51	0.51	0.5	0.48	0.35	0.21

tempo

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   35.48  103.98  120.05  121.29  131.03  207.97
```

Table 6: Changing the tempo

	100	125	175
Predicted Probability of Being Positive	0.46	0.49	0.56

duration

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   61385  189987  211667  218557  237987  478208
```

Table 7: Changing the Duration (ms)

	190000	215000	240000
Predicted Probability of Being Positive	0.54	0.49	0.45

Changes in rnb Model

```
##
## 0.6335431
```

Danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1400  0.5810  0.6930  0.6687  0.7750  0.9630
```

Table 8: Changing the Danceability

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.3	0.69	0.92

energy

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0118  0.4590  0.6000  0.5907  0.7250  0.9930
```

Table 9: Changing the Energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.43	0.82	0.96

loudness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -23.023  -9.656  -7.634  -8.106  -5.997  -0.882
```

Table 10: Changing the loudness

	-9	-6.5	-4
Predicted Probability of Being Positive	0.67	0.56	0.43

acousticness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000659 0.0456750 0.1705000 0.2680926 0.4312500 0.9890000
```

Table 11: Changing the acousticness

	0	0.04	0.2	0.5	1
Predicted Probability of Being Positive	0.54	0.56	0.61	0.71	0.83

Changes in rap Model

danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2110  0.6330  0.7360  0.7178  0.8200  0.9750
```

Table 12: Changing the Danceability

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.14	0.38	0.71

energy

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0161  0.5335  0.6600  0.6420  0.7720  0.9940
```

Table 13: Changing the Energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.24	0.49	0.75

key

```
##   0   1   2   3   4   5   6   7   8   9  10  11
## 245 449 215  55 164 161 209 236 227 168 222 236
```

Table 14: Changing the Key

	0	1	2	3	4	5	6	7	8	9	10	11
Predicted Probability of Being Positive	0.55	0.41	0.41	0.63	0.44	0.68	0.64	0.62	0.52	0.52	0.57	0.52

Loudness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -26.207  -8.690  -6.626  -7.212  -5.100   0.642
```

Table 15: Changing the Loudness

	-9	-6.5	-4
Predicted Probability of Being Positive	0.46	0.39	0.33

speechiness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0255  0.0702  0.1580  0.1854  0.2760  0.8650
```

Table 16: Changing the Speechiness

	0.05	0.1	0.25	0.5
Predicted Probability of Being Positive	0.3	0.34	0.47	0.69

acousticness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000022 0.0311000 0.1150000 0.2064776 0.3000000 0.9940000
```

Table 17: Changing the acousticness

	0	0.04	0.2	0.5	1
Predicted Probability of Being Positive	0.34	0.35	0.41	0.51	0.68

duration_ms

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   46000  168782  206187  209555  247182  461946
```

\begin{table}[H]

\caption{Changing the duration_ms}

	160000	210000	250000
Predicted Probability of Being Positive	0.36	0.41	0.45

\end{table}

Changes in rock Model

danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1160  0.4400  0.5430  0.5387  0.6440  0.9520
```

Table 18: Changing the Danceability

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.51	0.93	0.99

energy

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0167  0.5810  0.7540  0.7127  0.8760  0.9980
```

Table 19: Changing the energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.22	0.6	0.89

loudness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## -24.639  -9.969  -7.309  -7.838  -5.202   1.275
```

Table 20: Changing the loudness

	-9	-6.5	-4
Predicted Probability of Being Positive	0.71	0.52	0.33

mode

```
##    0    1
##  621 1465
```

Table 21: Changing the mode

	0	1
Predicted Probability of Being Positive	0.54	0.63

speechiness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02240 0.03202 0.03940 0.05591 0.05847 0.44200
```

Table 22: Changing the speechiness

	0.05	0.1	0.25	0.5
Predicted Probability of Being Positive	0.64	0.54	0.26	0.04

instrumentalness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000033 0.0002600 0.0655700 0.0130500 0.9690000
```

Table 23: Changing the instrumentalness

	0	0.05	0.1	0.5	1
Predicted Probability of Being Positive	0.65	0.63	0.61	0.43	0.23

tempo

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   37.11  105.01  123.95  124.01  139.98  207.97
```

Table 24: Changing the tempo

	100	125	175
Predicted Probability of Being Positive	0.56	0.63	0.74

duration

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   67400  207563  237367  247497  277237  513440
```

Table 25: Changing the duration

	160000	210000	250000
Predicted Probability of Being Positive	0.76	0.69	0.62

Changes in latin Model

```
##
## 0.816953
```

danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.0787  0.6660  0.7350  0.7192  0.7957  0.9790
```

Table 26: Changing the Danceability

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.4	0.79	0.96

energy

```
##     Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
## 0.000175 0.638000 0.745000 0.720501 0.837000 1.000000
```

Table 27: Changing the energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.49	0.79	0.94

speechiness

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.02320 0.04280 0.06535 0.09829 0.12200 0.59000
```

Table 28: Changing the speechiness

	0.02	0.08	0.25
Predicted Probability of Being Positive	0.79	0.81	0.86

acousticness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000493 0.0495250 0.1560000 0.2247010 0.3320000 0.9640000
```

Table 29: Changing the acousticness

	0	0.04	0.2	0.5	1
Predicted Probability of Being Positive	0.72	0.74	0.81	0.9	0.97

instrumentalness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000000 0.0000033 0.0517967 0.0004752 0.9940000
```

Table 30: Changing the instrumentalness

	0	0.05	0.1	0.5	1
Predicted Probability of Being Positive	0.83	0.82	0.8	0.64	0.39

Changes in EDM Model

```
##
## 0.1849687
```

danceability

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.1620  0.5710  0.6560  0.6523  0.7460  0.9830
```

Table 31: Changing the danceability

	0.5	0.7	0.75	0.8	0.85	0.9
Predicted Probability of Being Positive	0.03	0.3	0.44	0.6	0.74	0.85

energy

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  0.2490  0.7260  0.8400  0.8093  0.9180  0.9980
```

Table 32: Changing the energy

	0.5	0.7	0.9
Predicted Probability of Being Positive	0.1	0.15	0.22

acousticness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000054 0.0031250 0.0169000 0.0797360 0.0754500 0.9850000
```

Table 33: Changing the acousticness

	0	0.04	0.2	0.5	1
Predicted Probability of Being Positive	0.16	0.17	0.22	0.34	0.58

instrumentalness

```
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 0.0000000 0.0000437 0.0269000 0.2753124 0.6235000 0.9670000
```

Table 34: Changing the instrumentalness

	0	0.05	0.1	0.5	1
Predicted Probability of Being Positive	0.26	0.24	0.23	0.14	0.07

tempo

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   60.05  123.34  127.87  126.28  128.04  219.96
```

Table 35: Changing the tempo

	100	125	175
Predicted Probability of Being Positive	0.14	0.18	0.3