

Propuesta

Alberto, Ivan, Sara, Valeria

Propuesta de Proyecto Final

```
library(bayesrules)
library(tidyverse)
```

Introducción

De acuerdo con wikipedia, en la plataforma de Spotify se tiene acceso a más de 100 millones de canciones. Algunas canciones son más populares que otras. Si definimos a la variable Y como un rating de popularidad con $Y \in [0, 100]$. En general, entre más reproducciones recientes tenga una canción, su rating de popularidad será mayor.

Sabemos que la popularidad de una canción, no siempre es reflejo de su calidad, popularidad a largo plazo, o popularidad afuera de la audiencia de Spotify.

Además del artista hay otras características que hacen popular a una canción, como la “valencia”, que define que tan alegre suena una pieza.

Para el proyecto nos enfocaremos en entender:

- Cual es la popularidad típica de una canción en Spotify?
- Hasta donde influye el artista en la popularidad de la canción?
- Para un sólo artista, cómo varía la popularidad de sus canciones?
- Como influye la valencia en la popularidad de una canción?

Datos

Para este proyecto usaremos la base de datos “spotify” del paquete “bayesrules”.

```
data(spotify)
glimpse(spotify)
```

```
Rows: 350
Columns: 23
$ track_id      <chr> "7sYAS4CpbV90oSemgaEQat", "1YSwQvw1NrEPfA0j1iY8eV", ~
$ title         <chr> "On & On", "All The Lies", "Hear Me Now", "The Wall~
$ artist        <fct> Alok, Alok, Alok, Alok, Alok, Alok, Alok, Alok, Alo~
$ popularity    <dbl> 79, 56, 75, 65, 52, 45, 79, 61, 61, 61, 56, 71, 61, ~
$ album_id      <chr> "2a9AGivapFES2j2ElI3odn", "2FdyKqNeEX2H7gUhtlmJDC", ~
$ album_name     <chr> "On & On", "All The Lies (Remixes)", "Hear Me Now", ~
$ album_release_date <chr> "2019-11-22", "2019-07-19", "2016", "2019-06-28", "
$ playlist_name  <chr> "Dance Room", "Cardio", "ElectroPop 2020", "Electro~
$ playlist_id    <chr> "37i9dQZF1DX2ENAPP1Tyed", "37i9dQZF1DWSJHnPb1fOX3", ~
$ genre          <chr> "pop", "pop", "pop", "pop", "pop", "latin", "edm", ~
$ subgenre       <chr> "dance pop", "dance pop", "electropop", "electropop~
$ danceability   <dbl> 74.7, 70.7, 77.8, 68.2, 78.8, 73.5, 74.7, 68.2, 56.~
$ energy         <dbl> 77.6, 80.6, 46.3, 91.0, 43.0, 81.9, 77.6, 91.0, 90.~
$ key            <dbl> 10, 6, 11, 11, 11, 0, 10, 11, 11, 8, 1, 6, 11, 10, ~
$ loudness       <dbl> -5.898, -4.777, -7.603, -2.314, -7.757, -4.809, -5.~
$ mode           <dbl> 0, 1, 1, 1, 1, 1, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, ~
$ speechiness    <dbl> 7.22, 17.50, 3.89, 4.31, 4.19, 12.40, 7.22, 4.31, 3~
$ acousticness   <dbl> 2.3300, 0.9080, 54.6000, 0.9530, 53.7000, 37.9000, ~
$ instrumentalness <dbl> 1.38e-03, 0.00e+00, 2.89e-01, 6.73e+00, 2.61e-01, 0~
$ liveness       <dbl> 76.10, 36.60, 7.31, 7.32, 9.36, 4.23, 76.10, 7.32, ~
$ valence        <dbl> 67.00, 63.40, 49.60, 53.60, 46.10, 61.60, 67.00, 53~
$ tempo          <dbl> 125.070, 119.972, 121.999, 122.943, 121.996, 90.039~
$ duration_ms    <dbl> 147027, 163000, 194840, 159500, 192846, 186667, 147~
```

```
nlevels(spotify$artist)
```

```
[1] 44
```

La base de datos consta de 350 canciones de 44 artistas.

Métodos:

Para nuestro proyecto planeamos hacer tres tipos de modelos:

- **Modelo con parámetros homogéneos:** Modelar la popularidad de las canciones sin tomar en cuenta el artista.
- **Modelo con parámetros heterogéneos:** Modelar la popularidad de las canciones tomando en cuenta a los artistas que las interpretan.
- **Modelo Jerárquico** Una alternativa intermedia donde permitimos que la distribución inicial sobre la popularidad pueda adaptarse a los datos. Y comparar sus distribuciones predictivas con los datos observados, para ver cuál modela mejor y porqué.
- **Modelo Jerárquico con DAG:** Haremos un DAG para agregar al modelo jerárquico la variable de valencia.