



1ER TRABAJO PRÁCTICO DE DATA MINING

**Streams y características musicales en el top 200 Spotify
durante el año de la pandemia**

Carrasco, Lisandro; Opazo, F. Ayelén; Sotelo, Santiago

RESUMEN

El estudio llevado a cabo tuvo por objetivos explorar las características musicales presentes en las canciones que se posicionaron dentro del ranking 200 de Spotify durante el contexto de pandemia por Covid-19 en el año 2020; así como también conocer si existieron variaciones entre la cantidad de reproducciones (streams) de canciones que rankearon en general durante ese año con respecto a años anteriores. Por último, describir las características o features musicales de aquellas canciones que estuvieron en la posición top 1 en el año de la pandemia en comparación con el año previo. Los datos analizados fueron obtenidos de la API Spotify. Se encontraron indicios de que la posición en el ranking está más influenciada por factores tendenciales (modas) que por las características de los atributos musicales. Por otro lado, durante los primeros meses de declarada la pandemia no hubieron grandes picos de reproducciones; sin embargo, se encontró mayor cantidad de streams en el 2020, aunque también se halló cierta tendencia creciente entre 2018 y 2019 y, a su vez, una tendencia decreciente en valores de streams extremos considerados atípicos en el año de la pandemia con respecto a los años anteriores. Finalmente, los temas que ocuparon la primera posición del ranking durante 2020 mostraron cambios importantes en comparación a 2019 en sus características musicales, con descensos en variables que hacen a las buenas emociones que transmiten las canciones y aumentos en su energía, así como también un gran crecimiento de las versiones de recitales en vivo.

Palabras clave: pandemia, Spotify, features, streams, ranking

INTRODUCCIÓN

A finales del año 2019 en Wuhan, China, se detectaron 40 casos de neumonía de origen vírico. Esto llevó a la identificación de un nuevo tipo de coronavirus denominado Covid-19 (SARS-CoV-2). El 11 de marzo del año 2020, la Organización Mundial de la Salud (OMS) lo declaró como pandemia debido a su alto nivel de contagio, velocidad de expansión y gravedad. En un esfuerzo por detener la propagación del virus, los gobiernos de todo el mundo implementaron una serie de medidas preventivas, entre ellas el aislamiento social, un evento global disruptivo que generó cambios importantes en las actitudes y los consumos digitales de grandes segmentos de la población.

Dominio de aplicación y objetivos

Se quiere indagar, entonces, acerca de las características musicales (features) de las canciones más escuchadas del ranking principal 200 de Spotify durante el 2020 y observar si existió variación en la cantidad de reproducciones (streams), a partir de la situación mundial arriba descrita, con respecto a años anteriores. También, comparar en mayor profundidad features presentes en el top 1 del 2020 con el top 1 del año previo a la pandemia. El análisis del trabajo es guiado a partir de las siguientes preguntas:

1. ¿Qué características musicales poseen las canciones del top 200 de Spotify en el año de la pandemia?
2. ¿Hubo diferencias en la cantidad de reproducciones de canciones (streams) que rankearon durante el año 2020 a comparación de años anteriores?
3. ¿Existen variaciones entre las características presentes en las canciones del top 1 durante el año de la pandemia con respecto al año previo, 2019?

DATOS

Selección del dataset - Integración

Los datos seleccionados son extraídos a partir de un conjunto mayor de datos, obtenidos desde la API REST de Spotify en formato JSON y almacenados en una base de datos NoSQL MongoDB. Dentro de ese primer conjunto existen 3 colecciones, una para el almacenamiento de los nombres de artistas (*artist*) con 676 valores únicos, otra para los metadatos y características de las canciones que estuvieron en los rankings (*artist_audio_features_solo_art*) con 182.049 instancias, que contienen entre ellas documentos repetidos, y otra para la tabla de rankeos (*charts*) en un período que contempla desde enero de 2018 hasta marzo de 2021, con 63.600 instancias y presentando también documentos repetidos. Posteriormente también se recolectan nuevos datos complementarios a través del paquete de R '*spotifyr*'.

Siguiendo las preguntas planteadas para la propuesta del trabajo, se definen los siguientes criterios de corte y preselección de los datos a partir de estas 3 colecciones:

De ***charts***: crear vistas según el año al cual corresponden (2018, 2019 y 2020 - no se analiza 2021 a los fines del trabajo), ya que habrá instancias comparativas anuales, para poder trabajarlas

desde RStudio, quedando así 20.400 instancias para el año 2018 y 21.200 para los años 2019 y 2020. Luego, y ya desde RStudio, eliminar filas duplicadas con la función *unique()*, dejando un total de 10.200 observaciones para 2018, 10.400 para 2019 y 10.404 para 2020, por cada extremo filtrado. Por último, a propósito del primer objetivo, guardar un filtro de los data frames generados por las posiciones que se encuentran en valores del 1 al 20 en el año de la pandemia, generando data frames con canciones que estuvieron una o más veces dentro del top 20 a modo de hacer un zoom de las características que presentan, con sus respectivas cantidades de streams. Lo mismo se repitió para las posiciones entre los valores 1 y 10.

De ***artist_audio_features_solo_art***: este dataset aporta la información correspondiente a las características musicales de las canciones a partir de campos tales como *danceability*, *energy*, *loudness*, *speechiness*, *acousticness*, *instrumentalness*, *valence* y *tempo* y a los álbumes a los que corresponden las canciones. También se eliminan filas duplicadas, quedando 156.621 instancias. Se usan los metadatos que contiene para describir las canciones que componen los *charts* obtenidos, es decir, se hace un *left join* a través de la clave *url*, con limpieza previa de los datos de ambas tablas.

De ***artist***: estos datos se usan simplemente para cruzar con los otros dataframes y validar los nombres de artistas y las integraciones entre las distintas fuentes.

Recapitulando, los data frames integrados a preprocesar son los charts anuales y los charts del top 10 y 20 del 2020, juntados a sus respectivas características, asociadas por una columna *url* del track, con limpieza previa de los datos.

Eliminación de variables

Las variables que se mencionan a continuación se excluyen del análisis por considerarlas **irrelevantes o reiterativas** para los objetivos del trabajo:

De ***artist_audio_features***: *-album_id*, *-album_images*, *-album_release_date_precision*, *-analysis_url*, *-available_markets*, *-artists*, *disc_number*, *-key_name*, *-mode_name*, *-track_href*, *-track_number*, *-track_uri*, *-type*, *-is_local*, *-key_mode*, *-track_preview_url*, *-time_signature*, *-mode*, *-disc_number*).

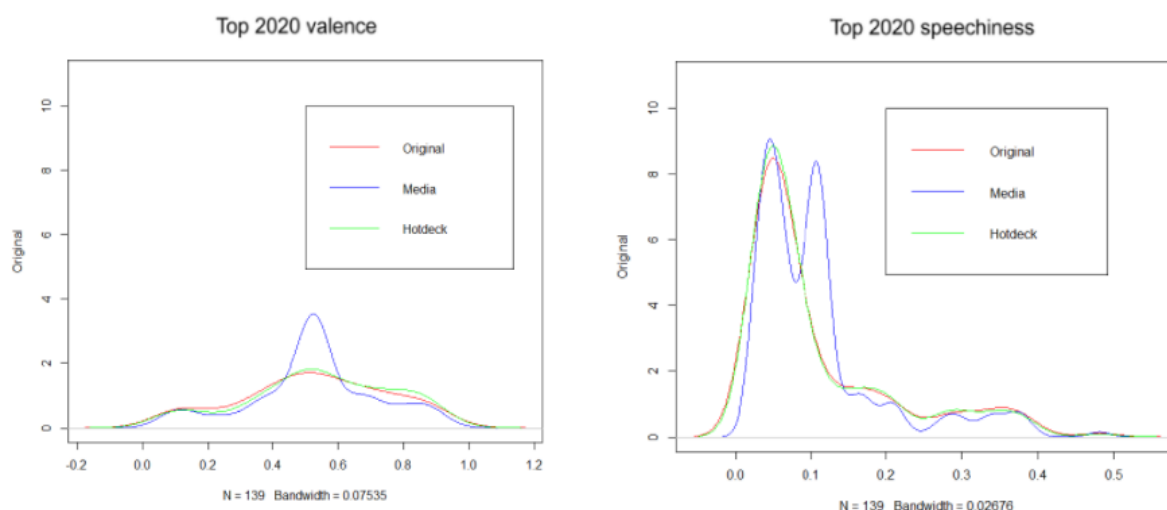
Por otro lado, por la forma en que Spotify mide el atributo *instrumentalness*, esta resulta ser una variable muy **ruidosa** y la mayoría de sus valores están concentrados en 0, presentando pocas observaciones relevantes. Este ruido la torna poco valiosa para el presente análisis y motivó que también sea eliminada. Se mantienen entonces 21 variables de un total inicial de 49 columnas correspondientes a ***artist_audio_features***.

Valores faltantes e imputaciones

A partir de las juntas de las tablas *charts* con ***artist_audio_features_solo_art*** se generaron **faltantes**, pese a haber realizado limpieza y adecuación de los datos a juntar previamente. En el año 2020, se observaron 30% de valores faltantes, mientras que en 2019 se presentaron un 21% y en 2018, un 38%. El 100% de los faltantes corresponden a variables originales de ***artist_audio_features_solo_art***, lo cual indica la presencia de tracks en el dataset de *charts* sin ninguna característica asociada en el otro.

Se prueban, entonces, distintas alternativas de tratamiento de estos datos faltantes para continuar el análisis. Por un lado, se evalúa la posibilidad de descartar las url de canciones que no

traen features de la otra tabla, para continuar con registros completos. También se comparan métodos de imputación puntual, específicamente sustitución por la media y **hot-deck**. Se muestran, a continuación, distribuciones originales de algunos de los features, filtrados por posiciones top 20 del año 2020, comparadas con las distribuciones utilizando los métodos indicados:



Vemos que la imputación por la media no es una opción viable ya que presenta un marcado sesgo por alejarse ampliamente de la distribución original y disminuir su varianza, justamente con tendencia hacia la media. Por otra parte, el método **hot-deck**, a simple vista, aproxima bastante bien a las distribuciones originales. Sin embargo, esta resolución implica la imputación de una proporción importante de datos y en prácticamente todos los features musicales de interés, teniendo así un peligro de sesgo considerable en la muestra en función de los datos presentes.

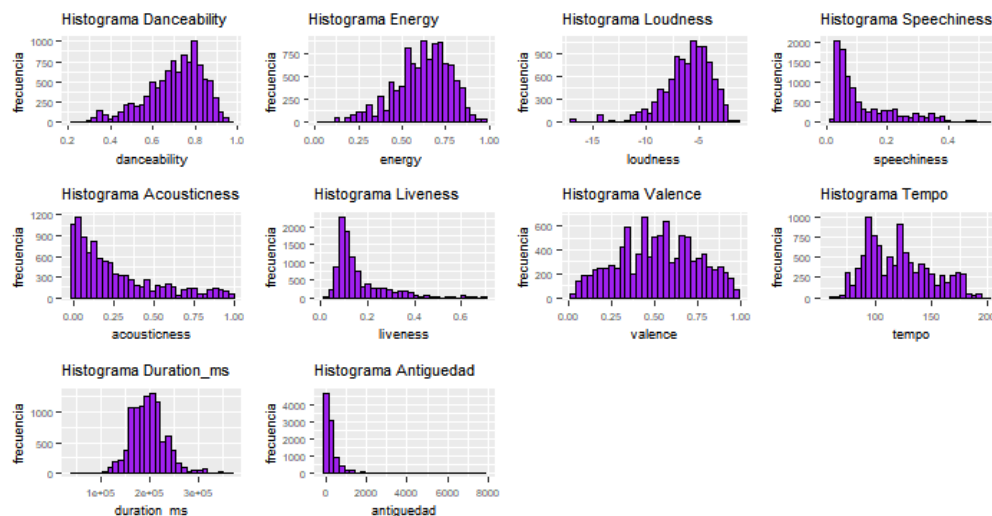
Ante esto, finalmente se opta por un método de imputación **cold-deck**, recurriendo al paquete de R, '**spotifyr**'. A través de la documentación de developers.spotify.com y su consola web, se comprende que las url de las canciones que presentan atributos faltantes no son reconocidos por la API de Spotify, pero sí se pueden tratar esas url como cadenas de caracteres y dividirlos de manera regular para desprender de ellos los *track_ids* de las canciones en cuestión, *track_ids* que sí son reconocidos por Spotify. Una vez detectado esto, se separan los datasets en dos, dejando por un lado los casos completos y por otro, los que presentan valores faltantes. En los datasets con valores faltantes se genera una nueva variable '**track_id**', producto de la división de url, y a partir de esa nueva información se logran descargar los atributos musicales de los temas mencionados con la función `get_track_audio_features()`. Lo mismo se repite para obtener la información faltante sobre los álbumes de estas canciones, con la función `get_tracks()`. Una vez descargados todos los temas únicos, el trabajo de **integración** consistió en identificar las columnas sobrantes, así como las que tenían la misma información con distinto nombre, para juntarlos nuevamente sin inconveniente. Finalmente, se obtienen datasets sin valores faltantes y con imputaciones de un marco de confianza importante, dado que el origen de estos datos imputados terminó siendo la propia API de Spotify.

RESULTADOS

1. Características musicales del top 200 en el año de la pandemia

Utilizando el set de información de las 200 canciones que componen los charts de 2020, se creó una variable adicional, “**antigüedad**”, como la diferencia entre la fecha de lanzamiento del álbum al cual pertenece el tema y la fecha de fin de semana a la cual corresponde la posición en el ranking. El propósito de agregar esta variable es sumarle valor al dataset original y descubrir si existe algún motivo adicional a los features originales que explique la posición de cada canción en el ranking. A continuación se observan las distribuciones de las variables:

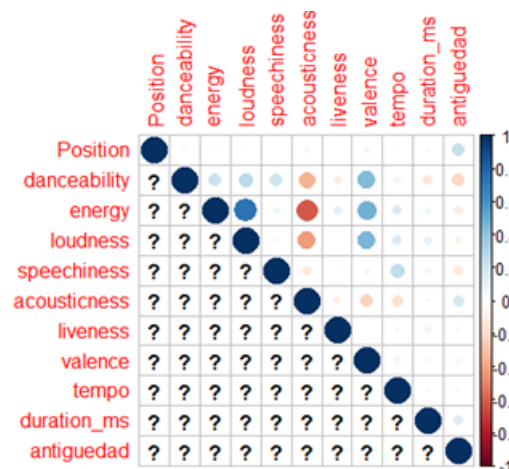
Distribucion de los atributos musicales en 2020



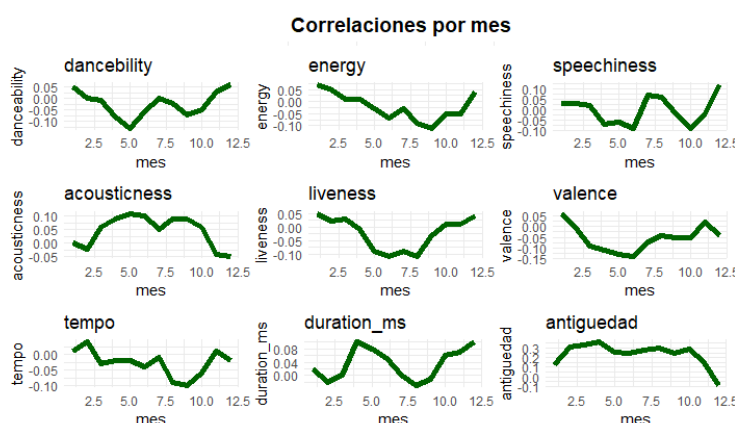
Se observa asimetría negativa en *danceability*, *energy* y *loudness*, asimetría positiva en *speechiness*, *acousticness*, *liveness*, *antigüedad* de forma fuerte y en *tempo* de forma suave. Hay una leve simetría en las distribuciones de *valence* y *duration_ms*.

Luego se calcularon las **correlaciones** entre los atributos musicales y la posición en el ranking.

Lo más interesante a destacar es que, a excepción de un valor moderado de antigüedad, no se observa correlación lineal relevante entre los atributos musicales y su posición en el ranking. La leve correlación negativa con antigüedad tiene sentido, ya que es esperable que las posiciones más altas del ranking sean ocupadas por temas más novedosos y por lo tanto de menor antigüedad (corroborado también por la asimetría positiva y su poca dispersión en valores bajos), pero también que dentro del universo de canciones disponibles haya cierta homogeneidad en la



variable antigüedad dentro del subconjunto de canciones del top 200. Luego de analizar las correlaciones anuales, se analizó su comportamiento durante el año para ver si existe divergencia significativa respecto de los valores anualizados.

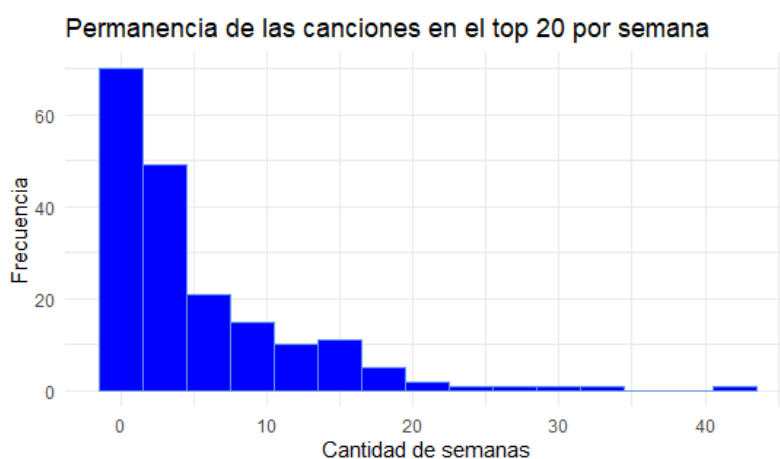


Se observa que, a excepción de *acousticness* y *antigüedad*, las demás correlaciones se comportaron de una manera levemente similar con respecto a la posición a lo largo del año. No se observan patrones significativos y las correlaciones son despreciables por su magnitud.

Permanencia en el Ranking

El análisis anterior arrojó indicios para suponer que las posiciones que ocupan las canciones en el ranking están vinculadas a un efecto tendencia (moda) más que a una combinación de atributos musicales. Para poder representar este efecto, se decidió crear una nueva variable adicional, *Permanencia en top 20*, tomando a ésta como el tiempo en semanas que una canción estuvo en ese rango del ranking. En base a esta premisa, vemos que el tiempo promedio de permanencia de una canción en el top 20 es de 5.8 semanas y la mediana es de 3 semanas. Además su correlación con la posición en el ranking es -0.37, indicando que tiene más peso el efecto tendencia para ir a posiciones cada vez más altas (*position* más bajo) que los valores de los atributos musicales para determinar el ranking de una canción.

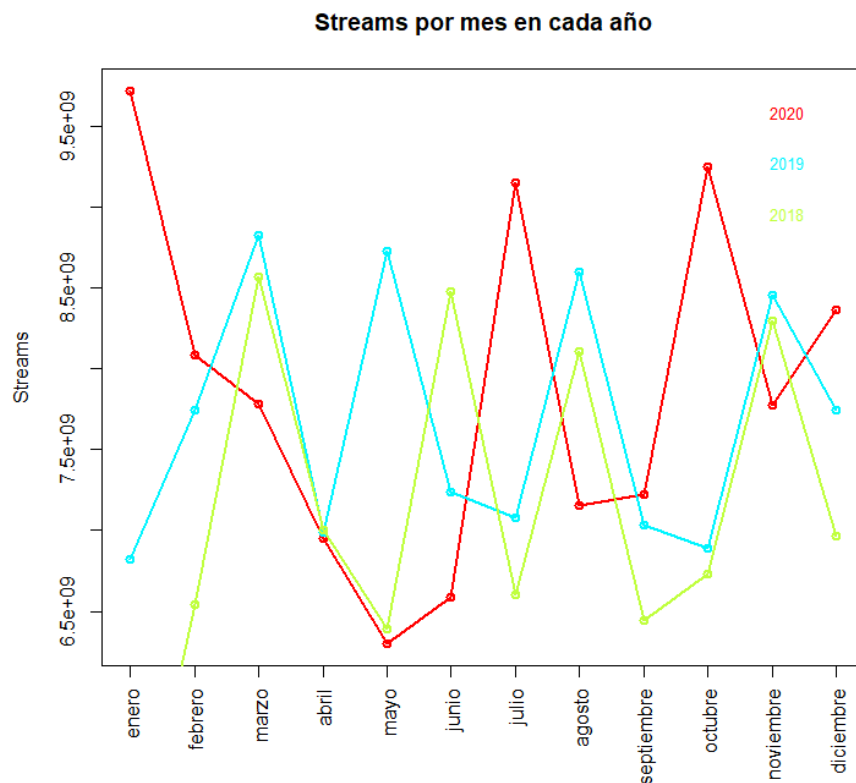
La distribución de la variable de permanencia está lejos de ser normal y presenta una asimetría positiva.



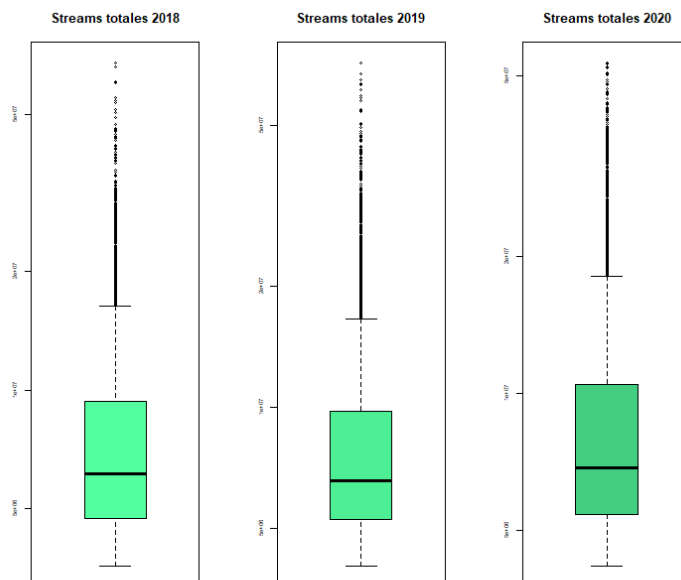
2. Comparación de cantidad de streams en 2020 y años anteriores

Top 200

Durante el año de la pandemia, las canciones que integraron el top 200 semanal de Spotify registraron un total de 94.904.089.319 reproducciones en la plataforma (mean=9.121.885; sd=6.113.369), un 5,2% más que en 2019 (mean=8.671.672; sd=5.863.022) y un 14,2% más que en 2018 (mean=8.148.149; sd=5.556.423). Veamos cómo fue esta distribución mensualmente:



Durante el 2020, puede apreciarse el pico más alto de reproducciones en el primer mes, previo a declararse formalmente la pandemia desde la OMS. Después la cantidad de streams fue disminuyendo hasta el mínimo de ese año, que fue en mayo. Luego existieron 2 picos más en julio y octubre, ya atravesando el contexto de pandemia. Con relación a años anteriores se ven los mínimos de reproducciones en enero, en oposición al 2020 (se destaca que en enero '18 los datos comienzan el día 12 de ese mes). Por otro lado, en abril parece haber una gran semejanza de streams para los 3 años. En mayo, que fue el mes de menor cantidad de reproducciones en el año de la pandemia, y también fue bajo en 2018, hubo contrariamente gran cantidad de reproducciones en 2019. De febrero a junio del año de la pandemia, la cantidad total de streams por mes estuvo por debajo de los totales de reproducción de ese mismo período para años anteriores, al igual que en agosto y noviembre. **2018 y 2019 parecen tener un comportamiento más similar entre sí en comparación al 2020**, aunque 2019 parece ir siempre un poco más arriba que su año anterior. Las grandes cantidades de streams del **año de la pandemia en los meses de picos, y en promedio, superan a las reproducciones de los años anteriores.**



En el gráfico de boxplots anterior se observan valores de streams mínimos similares entre los 3 años y típicos en cada distribución; sin embargo, el rango intercuartil del 2020 es más amplio que en años anteriores, extendiendo también su bigote superior y considerando menos valores atípicos en ese extremo.

Top 200

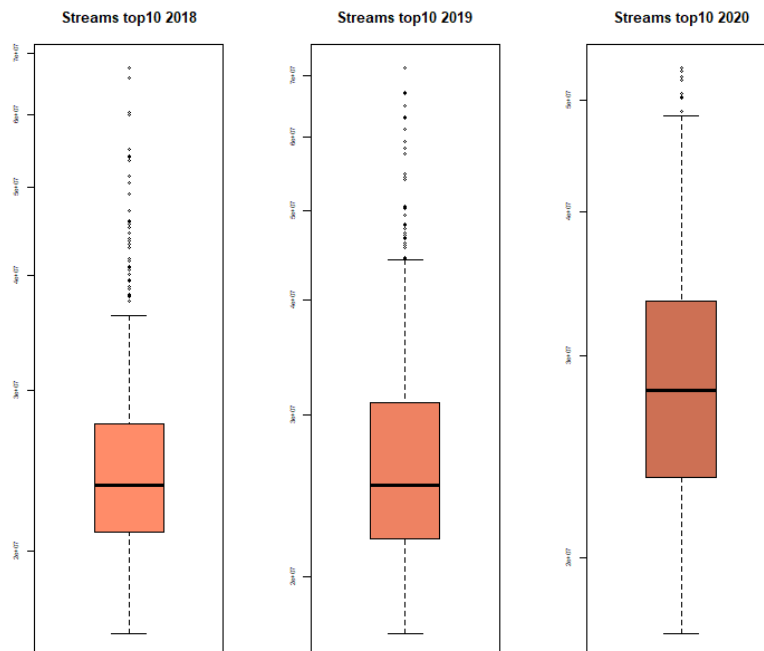
Análisis de valores atípicos

Si bien los 3 años muestran outliers sobre sus límites superiores (824 en 2018, 858 en 2019 y 799 en 2020), en proporción fueron disminuyendo hacia el año de pandemia, a la vez que se ampliaron los valores máximos de reproducciones, contabilizando un 8,1% de cantidad de reproducciones semanales que se consideran atípicas en 2018 por superar los 16.346.438 de streams, un 8,2% en 2019 que superaron los 16.532.007 de streams y un 7,7% en 2020 que superaron los 18.052.083 de streams. Es decir, en 2020 se registró una menor cantidad de outliers y, a su vez, una mayor cantidad máxima de streams. Para 2018, 155 tracks tomaron esos valores atípicos; para 2019, los tracks fueron 146; y para 2020, 145.

Estos tracks ocuparon las primeras posiciones para los 3 años, que van desde la posición número 1 hasta la 30. En cuanto a las semanas y meses de cada año, no hubieron concentraciones especiales (que podrían pensarse en los picos mensuales), sino que estos outliers se fueron produciendo durante cada año de principio a fin.

Top 10

Para las canciones que se posicionaron en el top 10 durante año de la pandemia, se produjeron un total de 15.200.338.105 de reproducciones (mean=29.231.419; sd=7.317.673), que representan un 16% de los streams totales de ese año en el chart 200. En relación a 2019, hubo un 5,4% más de streams totales dentro de este top (mean=27.741.353; sd=8.712.441) y un 16,2% más que en 2018 (mean=25.655.040; sd=7.483.557).

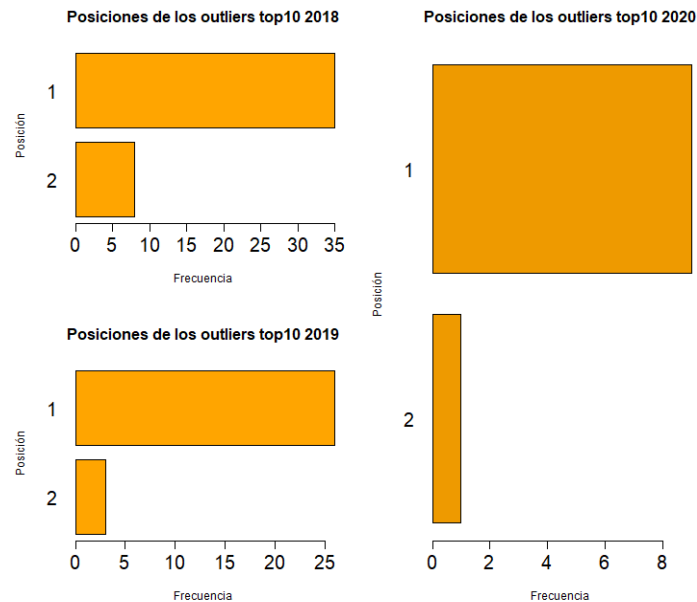


Al igual que para las distribuciones totales del top 200, los valores mínimos del top 10 son esperables y similares entre los años analizados y se presenta un orden creciente del rango intercuartil entre 2018 y 2020, siendo evidente el desplazamiento hacia arriba de la caja de 2020 y, por lo tanto, de su mediana con respecto a los años anteriores. También su distribución en el boxplot posee un bigote superior más largo, lo que genera un espectro más amplio para reconocer valores altos como esperables. **Si bien las medias fueron aumentando año a año, los valores máximos anuales crecieron entre 2018 y 2019 de 67.499.798 a 71.467.874 pero disminuyó en el 2020 a 53.401.383.**

Top 10

Análisis de valores atípicos

De hecho, en el top 10 se encuentra **una disminución en la cantidad de outliers año a año**, contemplando 43 observaciones que contienen valores atípicos de streams en 2018, 29 en 2019 y sólo 10 en el 2020. Porcentualmente, consistió en el 8,4% de streams en 2018 que estuvieron por encima del límite superior de 36.157.777, mientras que un 5,6% en 2019 estuvo por arriba de más de 44.148.365 streams y un 1,9% en 2020 tuvo más de 48.430.814 de reproducciones. Estos valores outliers se fueron dando a lo largo de cada año pero se concentraron entre las primeras 2 posiciones del ranking, principalmente de la 1:



14 tracks fueron considerados outliers al menos una vez en 2018 por su cantidad de reproducciones, mientras que 9 lo fueron para 2019 y sólo 5 para el 2020. En el siguiente gráfico se muestra cuáles fueron esas canciones y con qué frecuencia sus streams se consideraron valores extremos atípicos:



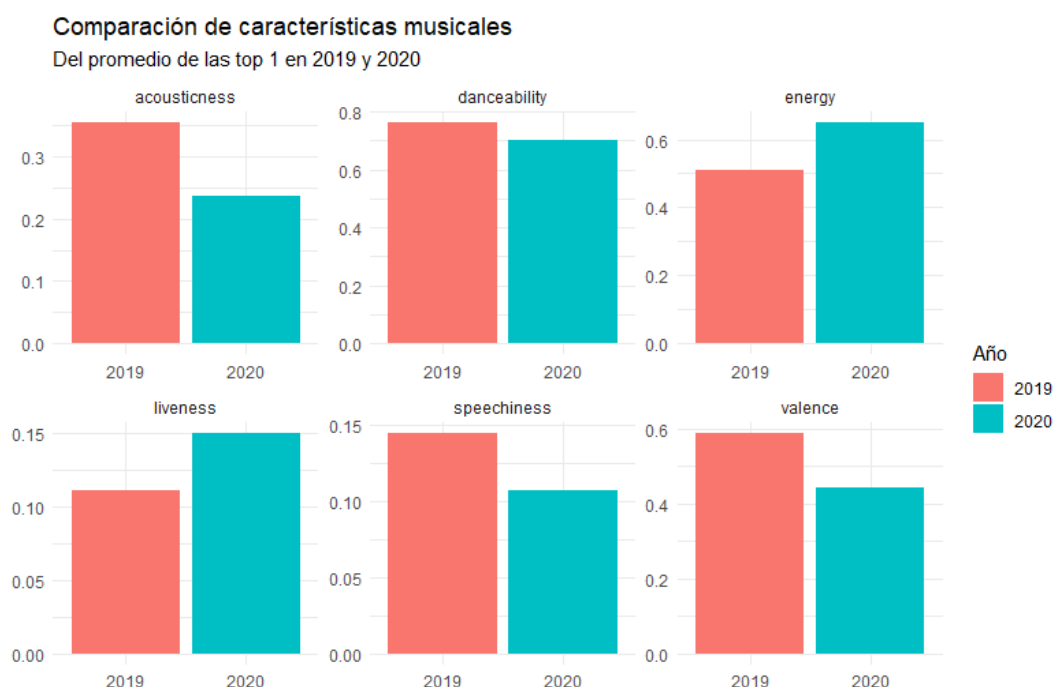
En línea con el desarrollo hasta el momento, vemos que en años anteriores a la pandemia hubo una mayor cantidad de outliers de extremo superior y que se presentaron con mayor frecuencia, en parte por haber un **corrimiento hacia arriba de valores medios de reproducción durante el año 2020**. 'God's plan', que además fue la moda del top 10 de 2018 en Spotify, fue el track que más

veces se reconoció como outlier, siendo 12 las veces que excedió el límite superior de streams. En el top 10 de 2019 la canción más frecuente fue 'Bad guy', sin embargo este tema sólo fue considerado outlier una vez ese año y el tema con mayor veces de reproducciones por encima del límite fue 'Señorita'. Algo similar ocurrió en 2020 con 'Blinding lights', que fue la moda del top 10 y también fue outlier una vez, pero el track con mayor cantidad de streams por encima del límite fue 'Dance monkey', que también fue outlier a finales del 2019, cuando fue lanzado el tema. Cabe destacar que la canción 'All I want for christmas is you' fue outlier del top 10 los 3 años seguidos pero solo una vez en cada uno.

3. Características musicales del top 1 del año de pandemia y del año anterior

Como último nivel de análisis, se realizó una comparación entre un segmento muy reducido, pero importante simbólicamente: las canciones que llegaron a ocupar el top 1 semanal entre 2019 y 2020. Se trata de 9 canciones en 2019 y 13 canciones en 2020 que han logrado encabezar el ranking y sostenerse reiteradas veces en la cima. En total, han generado 2.439.454.521 de reproducciones en 2019 (un 10,4% del total de reproducciones del chart anual) y 2.200.232.842 en 2020 (representando un proporción menor del total de la muestra, siendo un 8,88% de las reproducciones del chart de ese año), un descenso de casi el 10% que representa una pérdida de peso relativo del top 1 respecto de todo el ranking y una mayor dispersión de los temas que ocuparon esa posición.

Más allá de su aspecto masivo y la importancia de sus reproducciones, se observan las **características musicales** que las distinguen, entendiendo que ellas pueden expresar parte del estado de ánimo colectivo de la sociedad en el marco de incertidumbre que genera la pandemia.



Este gráfico de barras facetado compara la media de las características musicales del tema que ocupó la primera posición en cada año. Para llegar a este resultado, se promediaron los valores

de las 52 observaciones semanales, porque si bien esto implica la reiteración irregular de la presencia de todos los temas en cuestión, se entiende que esto resulta más representativo de la línea de tiempo analizada y el peso que ha tenido cada tema a lo largo de cada año. De su comparación, surgen algunas diferencias importantes. En promedio, los temas más de 2020 fueron:

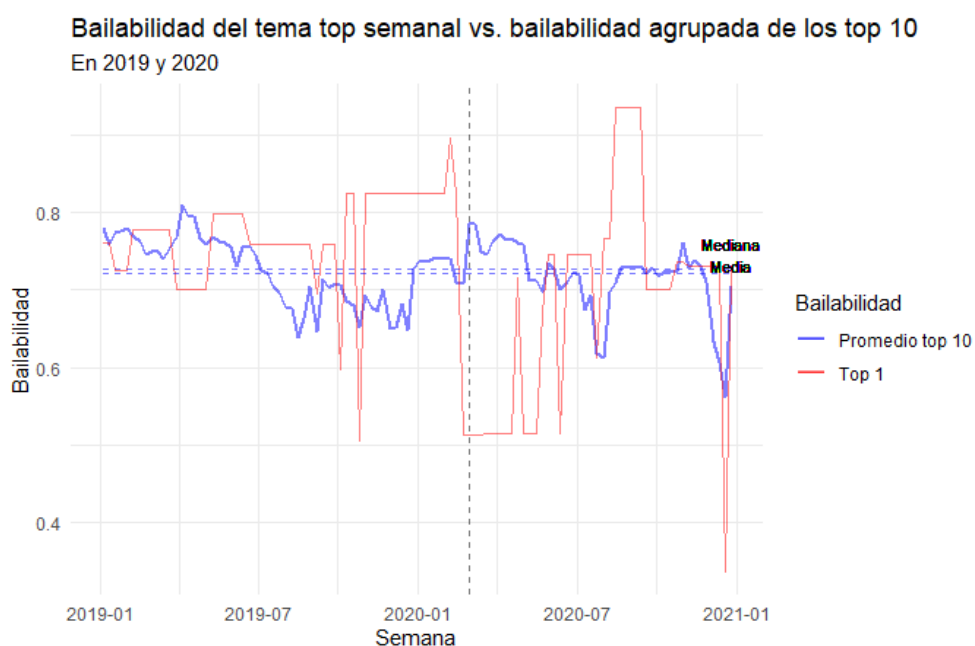
- Un 33% menos acústicos que los de 2019.
- Un 26% menos hablados.
- Un 25% menos alegres (*'valence'*).
- Un 8% menos bailables.
- Un 27% más enérgicos.
- Un 35% más en vivo.

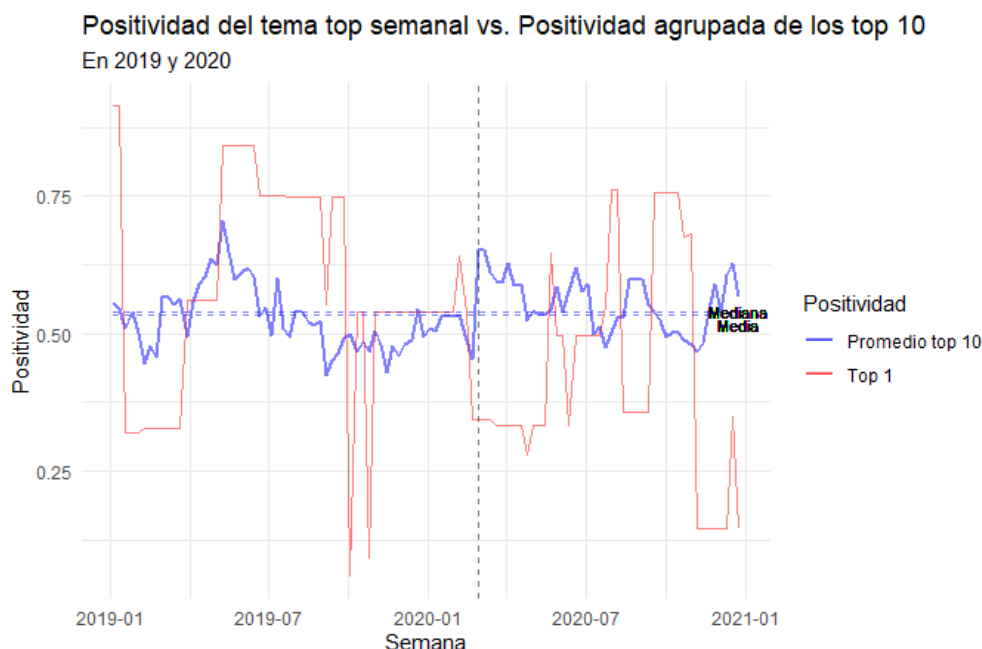
Estas variaciones en las principales características musicales entre un año y otro pueden resultar indicadores interesantes de un estado de ánimo colectivo menos optimista, reflejado en la caída de la variable que Spotify define como *'valence'*, una medida de la positividad y la alegría que expresa una determinada canción, así como en la caída de la bailabilidad de los temas. El aumento del promedio de la energía de cada tema también es una observación interesante. Spotify no define la energía de un tema como algo necesariamente positivo, si no que es una medida de su intensidad tiene una medida más baja.

Pero el principal efecto potencialmente asociable a los confinamientos y a las restricciones provocadas por la cuarentena parece ser el aumento de las reproducciones de los temas en vivo. Estos han crecido un 35% entre las canciones que ocuparon el primer lugar del ranking, insinuando una posible añoranza de las épocas en las que se realizaban conciertos musicales con público presente físicamente.

Evolución temporal

Los cambios interanuales pueden apreciarse en mayor detalle a partir de un análisis de la evolución semanal de las características de los temas que lideraron el ranking.



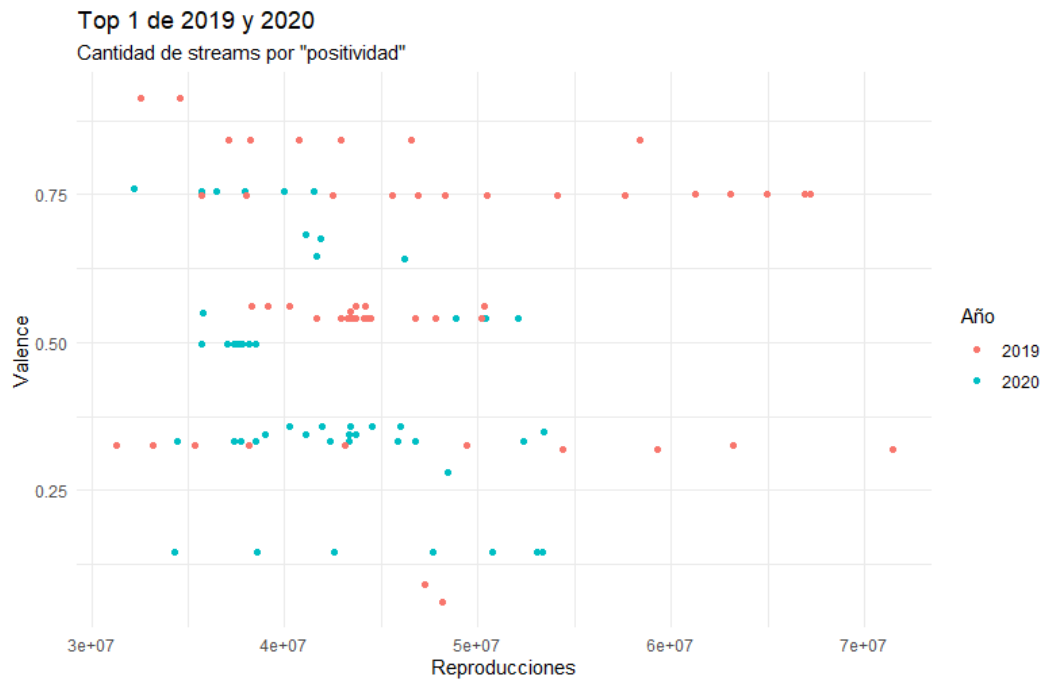


Los gráficos comparan las características musicales correspondientes al tema que encabezó el ranking en cada semana con el promedio de los temas que integraron el top 10 ese mismo período. Esto permite apreciar una evolución en el tiempo y hacer comparaciones en relación a la media y a la mediana del total de temas.

Puede verse con claridad que **a partir del 1ero de marzo de 2020** (línea rayada vertical), comienza una tendencia de **descenso de las variables musicales de positividad y bailabilidad, dos de las características que más se pueden asociar al estado de ánimo de los oyentes**. En prácticamente todas las semanas analizadas desde el primer momento pandémico importante en occidente, la bailabilidad y la positividad del tema que ocupó el primer lugar ha estado por debajo del promedio del top 10 de esa misma semana, así como ha estado por debajo del promedio general de todo el período en cuestión. Las medias del top 10 también presentan una tendencia descendente desde marzo del 2020.

Un agrupamiento anual

Por otro lado, un gráfico de puntos permite apreciar los cambios que se dieron en la relación entre los temas que ocuparon el primer lugar del ranking en ambos años y la variable ‘valence’. Se analiza esta relación a partir del gráfico de puntos por la importancia del cambio que hubo en esta variable entre los temas tops de un año y del otro, así como por lo significativo que resulta medir el “consumo de felicidad” en el marco pandémico.



Cada punto expresado en el gráfico es un top 1 semanal (52 observaciones por año), considerando que muchos de estos puntos serán temas repetidos porque ocuparon más de una vez el primer puesto, pero lo que varía es la cantidad de reproducciones que obtuvieron en esa semana en cuestión, por eso pueden verse alineaciones de puntos sobre ejes horizontales.

El uso de las tres variables en el mismo gráfico expone dos cambios que ya han sido mencionados, pero que se visualizan con más claridad en este caso: los temas del top 1 de 2020 han generado muchas menos reproducciones que los de 2019 y su “positividad” es menor en la mayoría de los casos. Si bien no hay una delimitación estricta, el gráfico de puntos muestra un cierto grado de agrupamiento entre los registros de un año y de otro, siendo los del 2019 los que se ubican en la parte superior del gráfico y los del 2020 los que se posicionan debajo (divididos por la positividad), así como también permiten detectar posibles outliers, como el punto ubicado en el extremo derecho del eje de reproducciones, muy alejado de los otros registros de 2019.

DISCUSIÓN

- En el análisis del top 200, no se encontró evidencia de una relación entre las preferencias de los usuarios respecto de los atributos musicales de una canción y su posición en el ranking, pero hay indicios de una incidencia de la antigüedad del tema y de efectos tendenciales (modas) que impactan en su preferencia (ser o no escuchados).
- A pesar de la tendencia creciente de reproducciones anuales que se registra entre 2018 y 2020, una selección más reducida muestra que los temas que ocuparon las mejores posiciones del ranking durante 2020 han generado menos reproducciones que en otros años.
- El comienzo de los aislamientos preventivos por la pandemia se vio acompañado por una caída notable en las reproducciones en la plataforma de Spotify, generando en marzo un mínimo absoluto de streams acumulados en un mes.
- Esto puede tener una doble interpretación: por un lado, anímica, los usuarios de la plataforma podrían haber reaccionado al aislamiento con una reducción de su consumo musical producto de la preocupación e incertidumbre generada por el coronavirus; por otro lado, el consumo de Spotify está muy relacionado al tipo de movilidad principalmente urbana de sus usuarios, su consumo en el transporte o en desplazamientos de un lugar a otro se vio seriamente afectado producto del cese abrupto de viajes realizados por los consumidores de la plataforma.
- El pico de julio de 2020 podría tener una vinculación interesante con la reapertura desmedida que tuvo lugar durante el verano europeo, producto del descenso de casos de ese entonces y de la necesidad de mantener el mercado de turismo.
- Las características musicales de los temas más escuchados podrían resultar un interesante indicador demoscópico del estado de ánimo generalizado de la sociedad: a partir de estas observaciones se podrían intuir ciertos efectos psicológicos que hayan tenido el encierro y la pandemia. Principalmente, esto podría apreciarse en la tendencia decreciente de las características de positividad y bailabilidad de los temas más escuchados durante 2020.
- Al tiempo que el aumento de la presencia de canciones grabadas en vivo, son una señal fuerte de la añoranza de las épocas en que la presencialidad no suponía un peligro para la salud, un importante componente nostálgico.
- Algunas **limitaciones** del trabajo tienen que ver con el sesgo acotado a las canciones del top 200 de Spotify. Para futuras líneas de investigación sería interesante analizar características de una muestra de tracks más ampliada, así como también información correspondiente a los usuarios que utilizan la plataforma y observar de qué manera interactúan con la elección de estos tracks y sus características.

REFERENCIAS

- COVID-19: cronología de la actuación de la OMS (2020) en: [27 de abril de 2020 Declaración COVID-19: cronología de la actuación de la OMS](#)
- Jiawei Han, Micheline Kamber, Jian Pei. (2011). Tercera edición. Data Mining: Concepts and Techniques.
- [Web API Reference. Spotify for developers](#)
- [Exploring the Spotify API with R: A tutorial for beginners, by a beginner](#), Mia Smith