

Análisis de las canciones más escuchadas entre 2010 y 2019

Todos hemos escuchado alguna canción en una fiesta y pensado: ¿Cómo puede esta mi ** ser tan popular?

Por ello estamos hoy aquí; para entender alguna información interesante de la música famosa, pero sobre todo, para ver si existe una fórmula secreta para crear esa basura de música que escucha todo el mundo

Si no hay código debajo de este texto, estamos en una versión del análisis realizado en la que todo el código ha sido extraído para facilitar la lectura de personas no interesadas en él, sino en el análisis de los datos.

Para tener acceso al GitHub con el código original, hacer clic en el [enlace](#).

Guía

1. [Primer vistazo a los datos](#)
2. [Referencias](#)

1. Primer vistazo a los datos

Antes de nada, veamos una lista de las 10 canciones más populares de 2010 a 2019 según Spotify:

- 1° Memories - Maroon 5
- 2° Lose You To Love Me - Selena Gomez
- 3° Someone You Loved - Lewis Capaldi
- 4° Señorita - Shawn Mendes
- 5° How Do You Sleep? - Sam Smith
- 6° Trampoline (with ZAYN) - SHAED
- 7° South of the Border (feat. Camila Cabello & Cardi B) - Ed Sheeran
- 8° Truth Hurts - Lizzo
- 9° Good as Hell (feat. Ariana Grande) - Remix - Lizzo
- 10° Happier - Marshmello

La calidad de setas canciones queda a gusto del lector, pero creo que da un buen punto de partida para preguntarnos qué las hace tan populares

A continuación dejaremos algunas estadísticas acerca de 600 de las canciones más populares que pueden ser de interés.

Sin embargo, todavía no será fácil sacar ninguna conclusión, con las cuales empezaremos en el capítulo 2, mediante el uso de algunos algoritmos de aprendizaje automático

Se adjunta en [este pdf](#) el significado de cada dimensión de los datos que tenemos, cuyo nombre ha sido abreviado en este análisis para compactar algunas tablas y gráficos

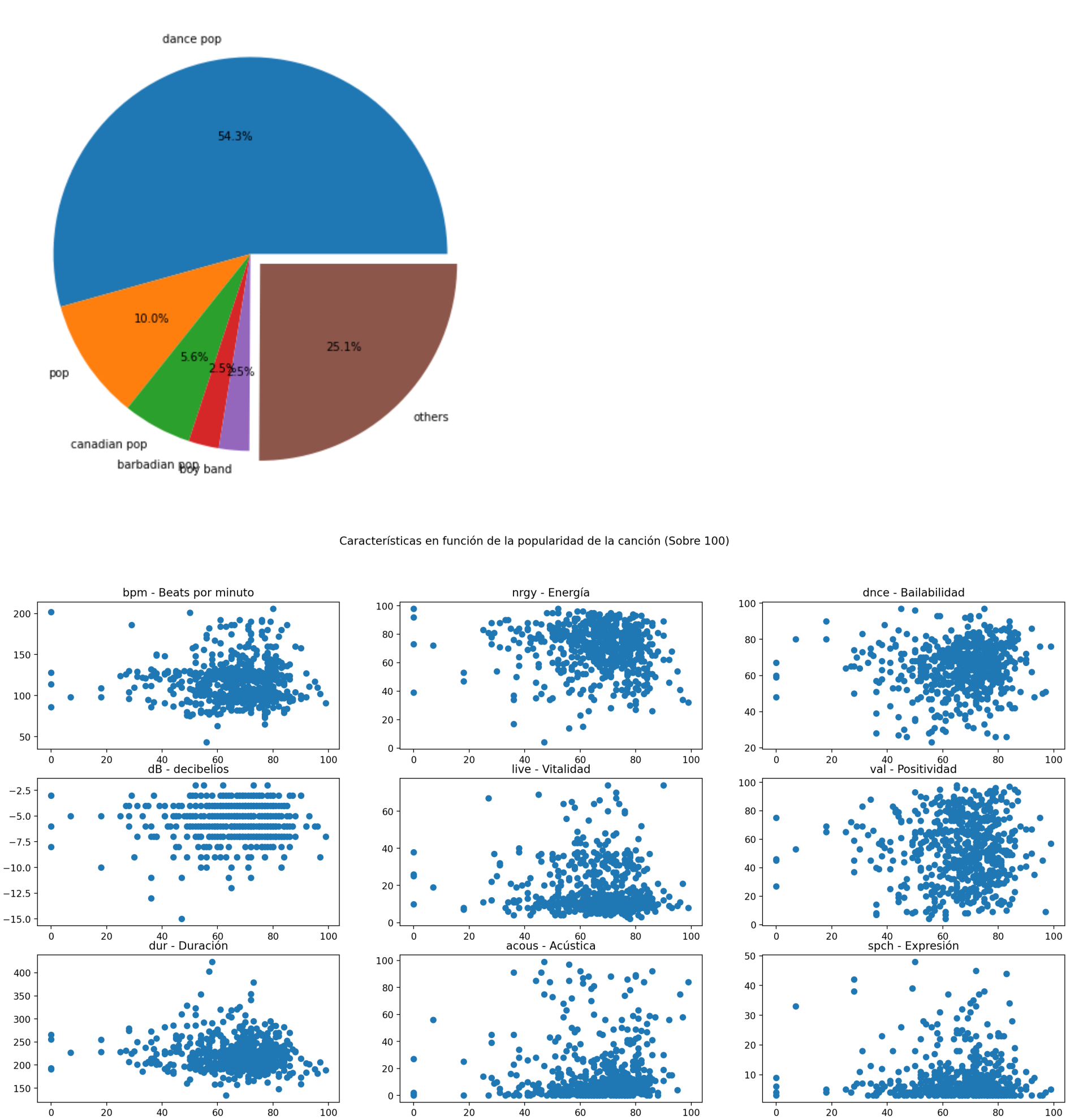
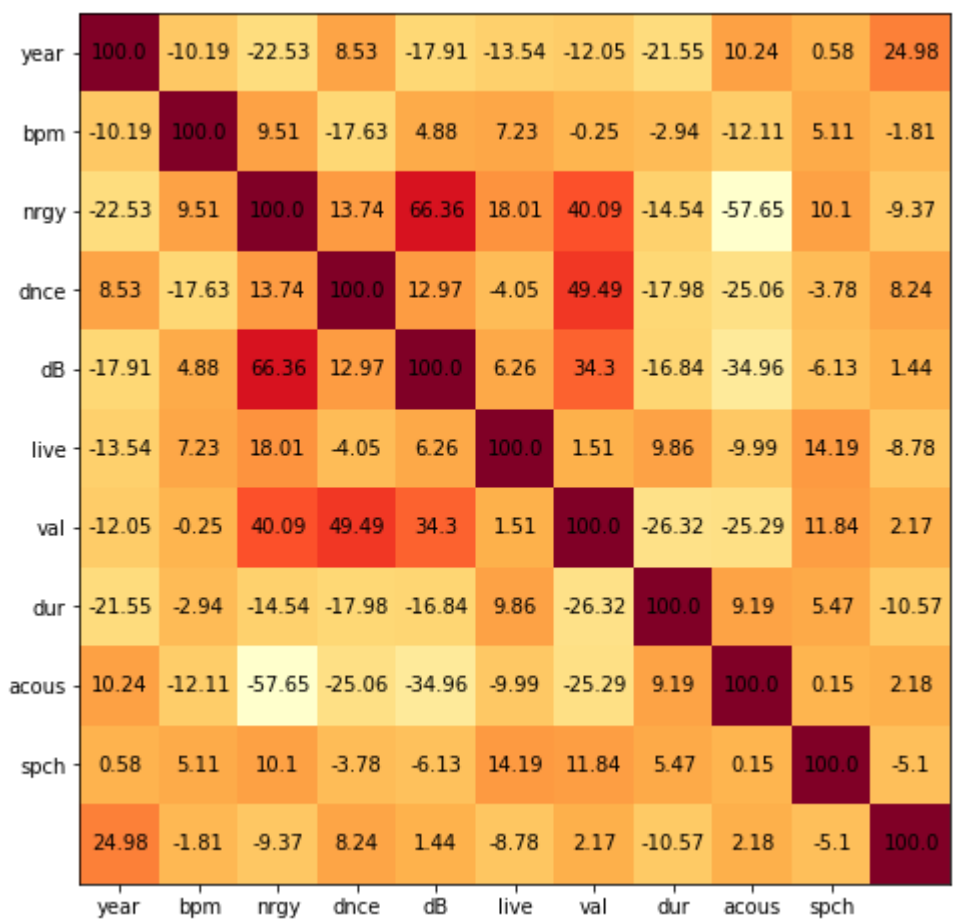


Tabla con la media, desviación típica, y cuantiles de cada característica que tenemos

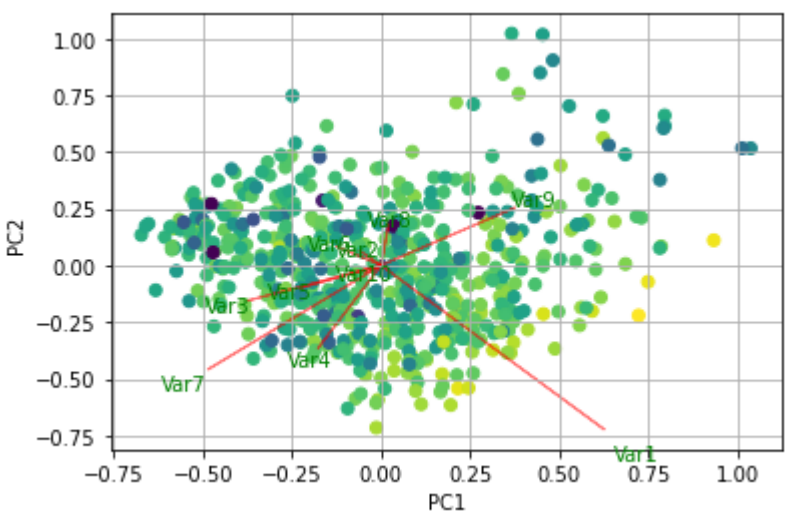
	count	mean	std	min	25%	50%	75%	max
year	602.0	2014.589701	2.608592	2010.0	2013.0	2015.0	2017.00	2019.0
bpm	602.0	118.742525	24.339497	43.0	100.0	120.0	129.00	206.0
nrngy	602.0	70.621262	16.068471	4.0	61.0	74.0	82.00	98.0
dnce	602.0	64.486711	13.129357	23.0	57.0	66.0	73.00	97.0
dB	602.0	-5.488372	1.704657	-15.0	-6.0	-5.0	-4.00	-2.0
live	602.0	17.803987	13.093347	2.0	9.0	12.0	24.00	74.0
val	602.0	52.312292	22.430640	4.0	35.0	52.0	69.00	98.0
dur	602.0	224.671096	34.158310	134.0	202.0	220.5	239.75	424.0
acous	602.0	14.350498	20.775203	0.0	2.0	6.0	17.00	99.0
spch	602.0	8.372093	7.481608	3.0	4.0	5.0	9.00	48.0
pop	602.0	66.631229	14.273775	0.0	60.0	69.0	76.00	99.0

2. PCA(Análisis de Componentes Principales)

Matriz de correlación porcentual de los valores que tenemos



Dado que el análisis mediante aprendizaje automático de grandes cantidades de datos es altmanente costoso, y este coste aumenta exponencialmente con varias dimensiones, antes de aplicarlo se suele realizar una reducción de dimensionalidad, creando unas nuevas dimensiones abstractas que maxifiquen la eficiencia del programa, y en este caso lo haremos mediante PCA Tal vez, con 600 canciones a analizar, no sea necesario el uso de este algoritmo, aunque siempre es una buena técnica para sacar más datos en el análisis exploratorio



99. Referencias

El archivo usado para el análisis ha sido extraído de la plataforma de bases de datos kaggle, que es open source, por lo que tenemos todo el derecho a hacer lo que queramos con los datos.

Adjunto link a la base de datos [aquí](#).