

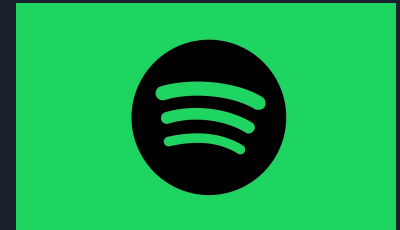
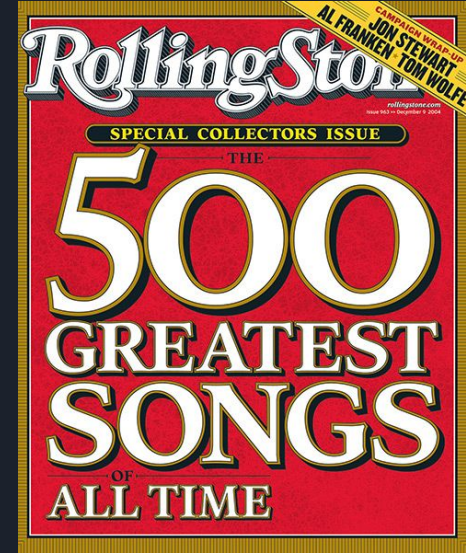


Análisis de dataset de Spotify y Top 500 canciones de Rolling Stones

Miguel Calvo
Alejandro Muñoz
Rodrigo Juárez

Bases de Datos

Teníamos curiosidad sobre qué hace que una canción sea una "gran canción". Para responder a esta pregunta trabajamos con dos conjuntos de datos. La primera es una base de datos de Spotify que encontramos en Kaggle que contiene más de 160.000 canciones con todo tipo de métricas sobre la canción (acústica, bailabilidad, energía, duración, instrumentalidad, valencia, popularidad, tempo, vivacidad, sonoridad, locución), la segunda son las 500 mejores canciones de todos los tiempos, según la revista Rolling Stones. Este segundo conjunto de datos únicamente contiene el nombre de la canción, el artista y el año de lanzamiento.





Unión de Base de datos por Hashes

Se propone el uso de MinHash para unir las canciones entre ambas bases, pues:

- # Comparaciones sin MinHash vs con MinHash : 21,281,102,250 vs 1,243,402
- Hallazgo de tejas similares

Preprocesamiento

- Rango de años entre 1948 y 2008
- Eliminación de canciones con caracteres que no pertenecen al alfabeto inglés
- Eliminación de palabras como “Sonata”, “Nocturne”, “A Major”, etc.
- Variable para calcular tejas: Título + “ ” + Artista, en minúsculas y sin caracteres especiales

Hiperparámetros

- $k = 4$
- # funciones hash = 4

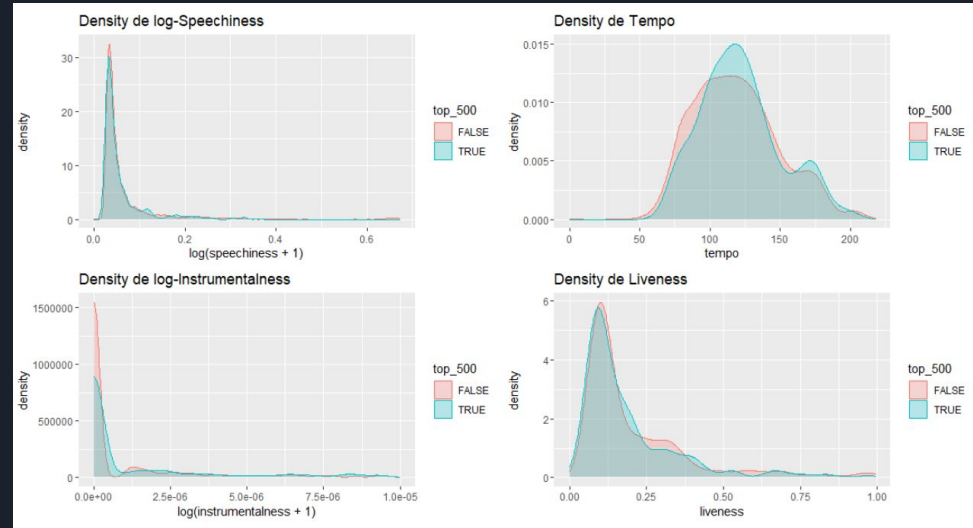
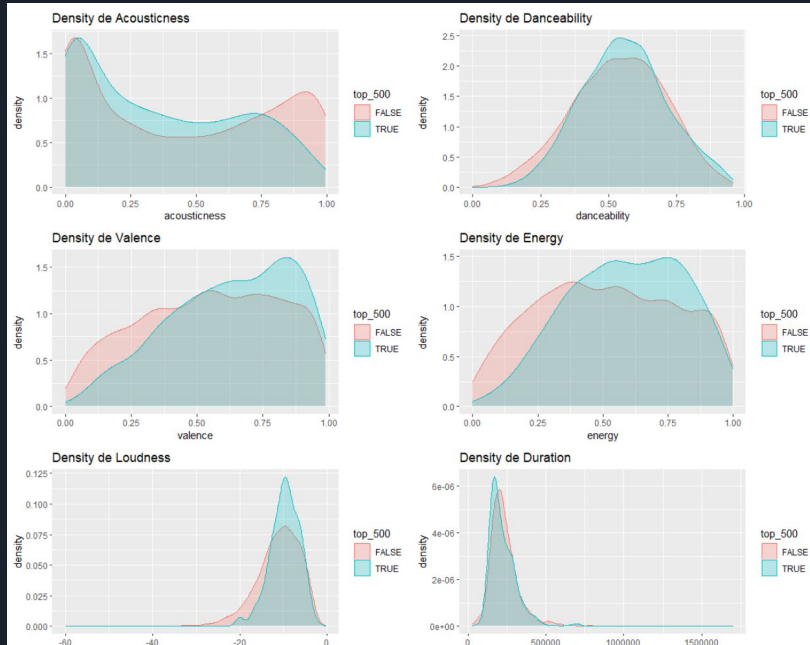
Resultados (score > 0.3)

- 440 hallazgos
- 30 falsos positivos

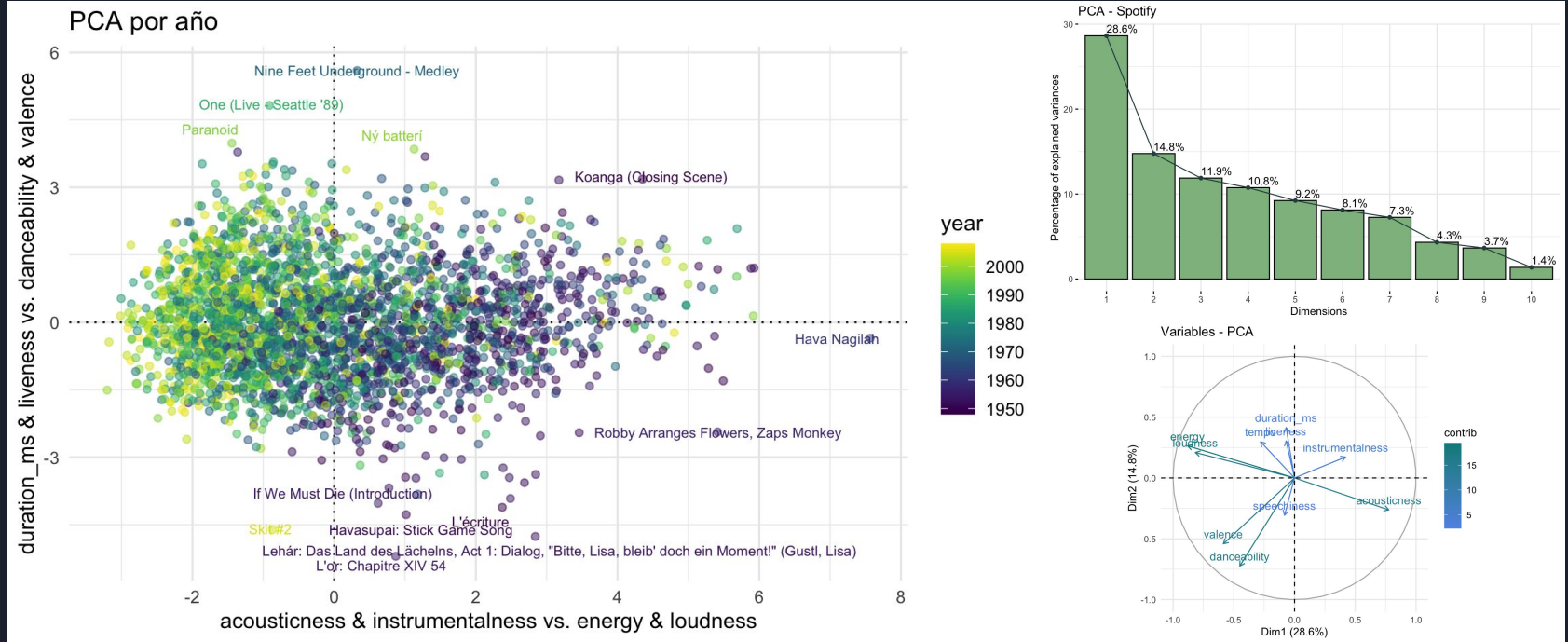
Unión de Base de datos por Hashes

Score	Rolling Stones - Title	Spotify - Title	Rolling Stones - Artist	Spotify - Artist
0.636	Gloria	Gloria	Them	['Them']
0.634	Good Lovin	Groovin'	The Young Rascals	['The Young Rascals']
0.633	My Sweet Lord	My Sweet Lord (2000) - 2014 Mix	George Harrison	['George Harrison']
0.632	96 Tears	96 Tears	? and the Mysterians	['? & The Mysterians']
0.632	The Sounds of Silence	The Sounds of Silence	Simon and Garfunkel	['Simon & Garfunkel']
0.632	Why Do Fools Fall In Love	Why Do Fools Fall in Love	Frankie Lymon and The Teenagers	['Frankie Lymon', 'die Teenagers']
0.630	Oh	Leah	Roy Orbison	['Roy Orbison']
0.627	We Gotta Get Out of This Place	We Gotta Get Out Of This Place (US Version)	The Animals	['The Animals']
0.625	What'd I Say	What'd I Say, Pt. 1 & 2	Ray Charles	['Ray Charles']
0.625	The Loco-Motion	The Locomotion	Little Eva	['Little Eva']
0.623	I Want to Know What Love Is	I Want to Know What Love Is - 1999 Remas	Foreigner	['Foreigner']

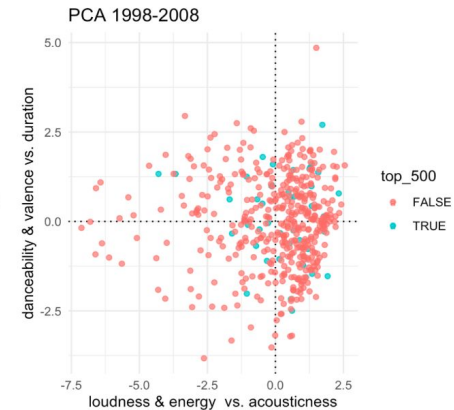
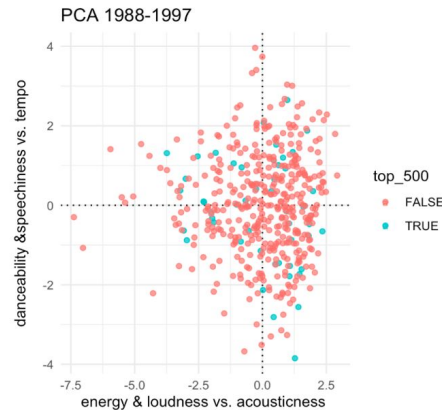
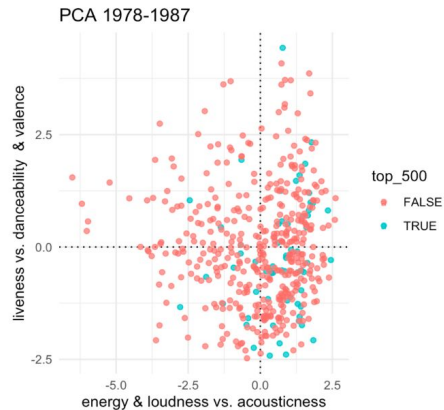
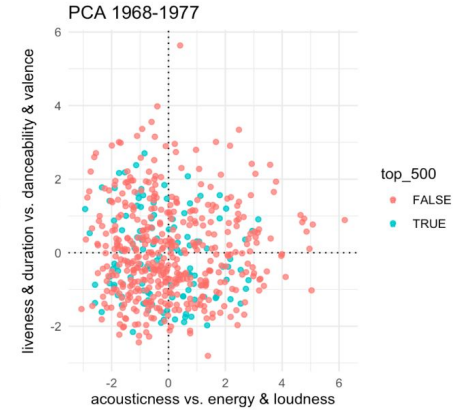
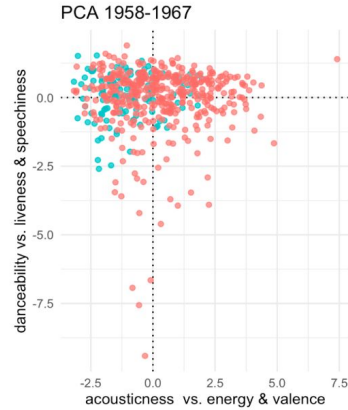
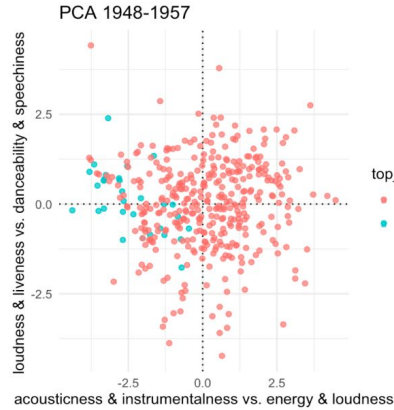
Comparación de variables (Top 500 y Spotify)



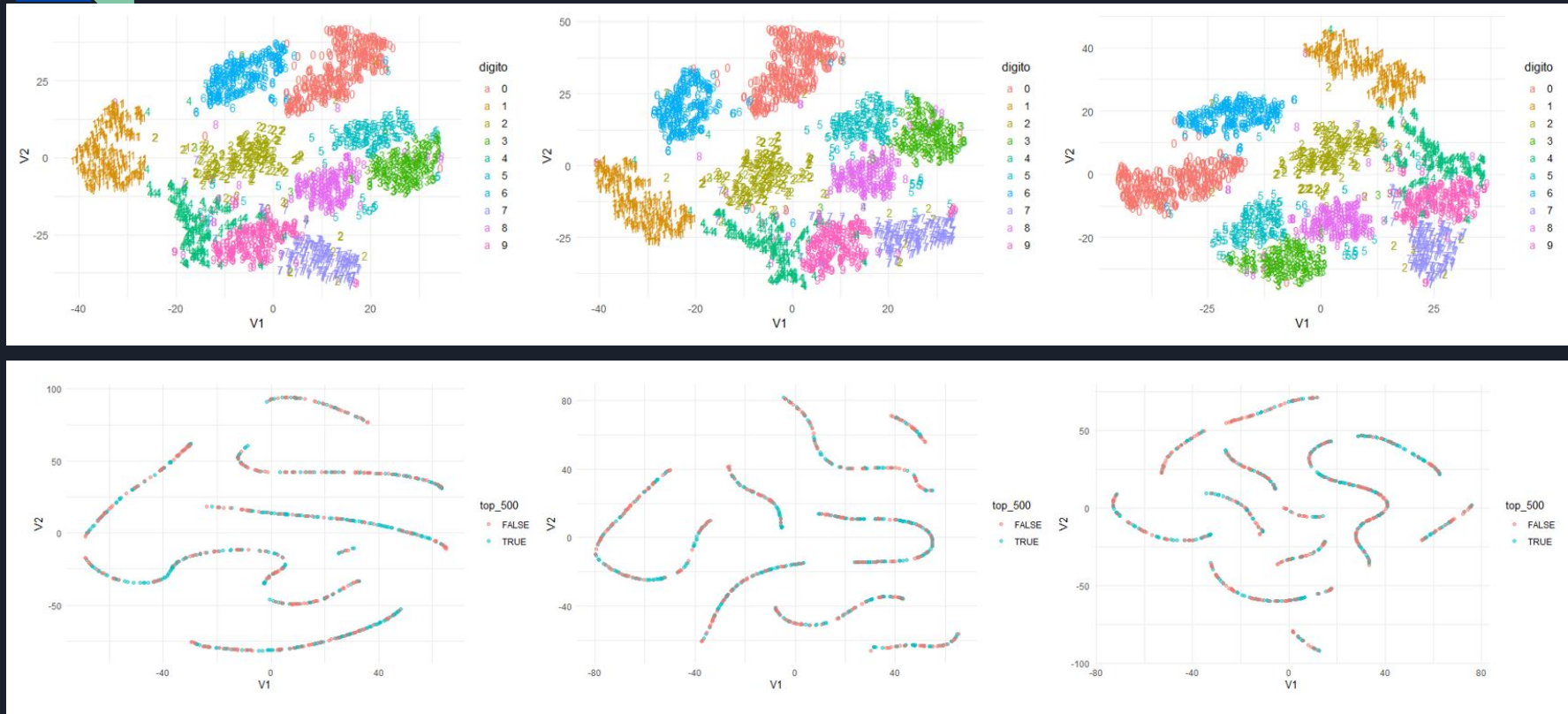
Análisis de Componentes Principales



Análisis de Componentes Principales



TSNE



Conclusiones

- Join por Hashes
- PCA
- TSNE

