

Alejandro Chaires

Spotify Classification



Contexto y objetivos

Con el proposito de retener y obtener nuevos clientes, Spotify debde de tener una gran seleccion de generos musicales, tambien debe de contar con una interfaz sencilla y facil de utilizar. Pero a su vez necesita contar con un sistema de recomendaciones de contenido. Esto para lograr que usuarios puedan conocer mas artistas, canciones y contenido en la plataforma. Esto ayuda a que podamos hacer crecer los usuarios en la plataforma.

Es aqui donde nace la idea de crear un sistema de recomendacion de contenido y clasificacion en base a gustos de los usuarios del servicio. Todo esto para poder dar visibilidad a canciones recien ingresadas al sistema y poder dar presencia a nuevos temas.

Este sistema se basara en las variables y relaciones entre cada tema con mayor gusto, asi como diferentes datos que son mostrados en el dataset a analizar.

Preguntas de interés

¿Hay diferencias estadísticas significativas entre las variables que clasifican las canciones?

¿Cuáles son las correlaciones en los diferentes parámetros?

¿Cuáles son los géneros más populares?

¿Cuáles son los artistas más populares?

¿En qué variable se destacan las canciones más populares?

¿Qué porcentaje de las canciones tienen letras explícitas?

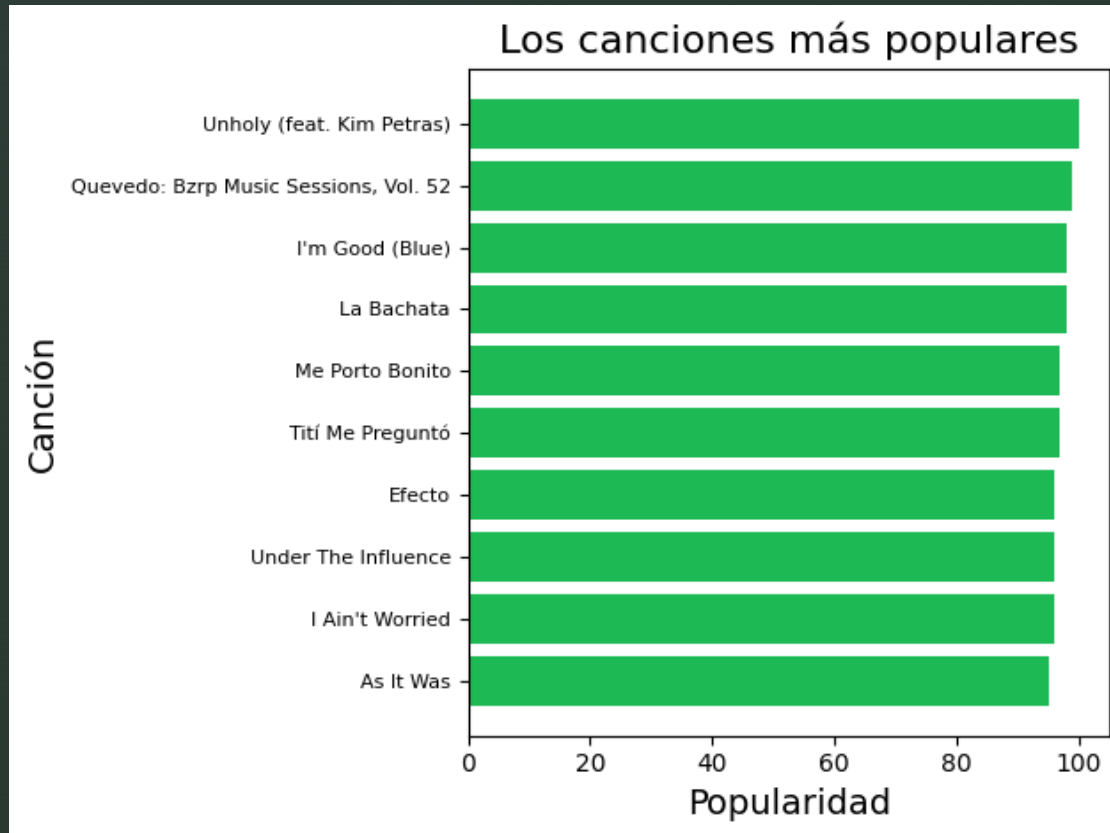
¿Cuál es la duración promedio de las canciones?

Popularidad



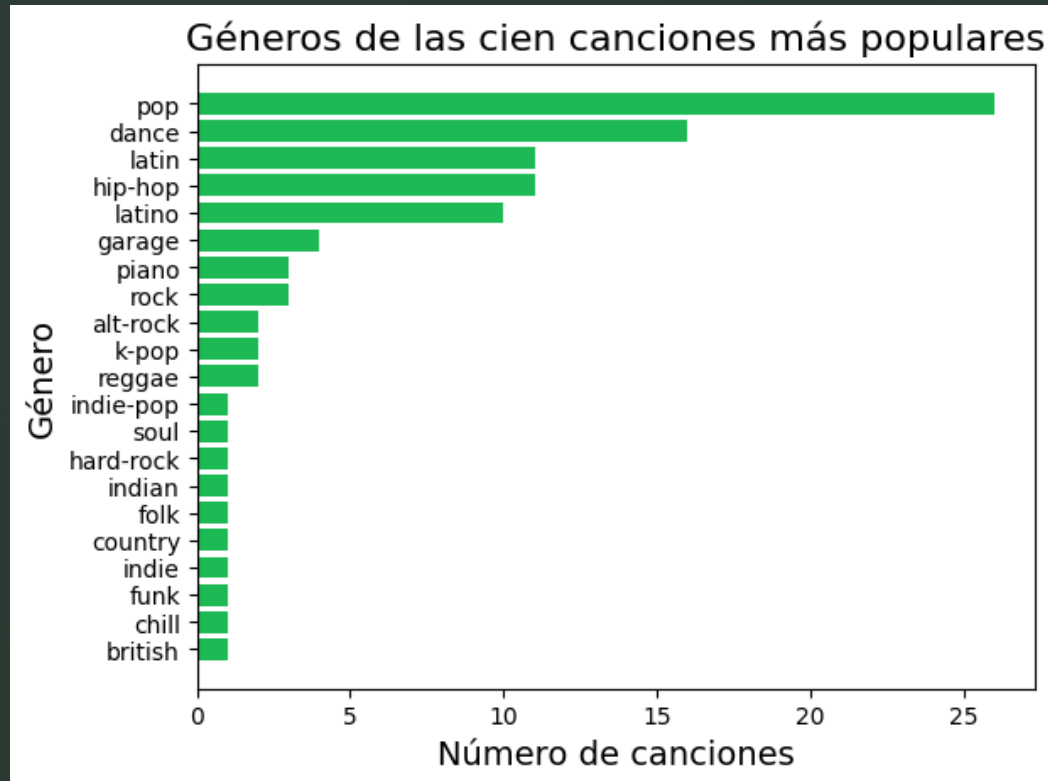
El Word Cloud muestra los 25 artistas más populares del dataset original, la mayoría de ellos son artistas vigentes y de gran difusión en la actualidad. Esto ilustra en gran medida los gustos de los oyentes, dando una idea aproximada de los artistas que más serán solicitados al algoritmo de recomendación de canciones.

Popularidad



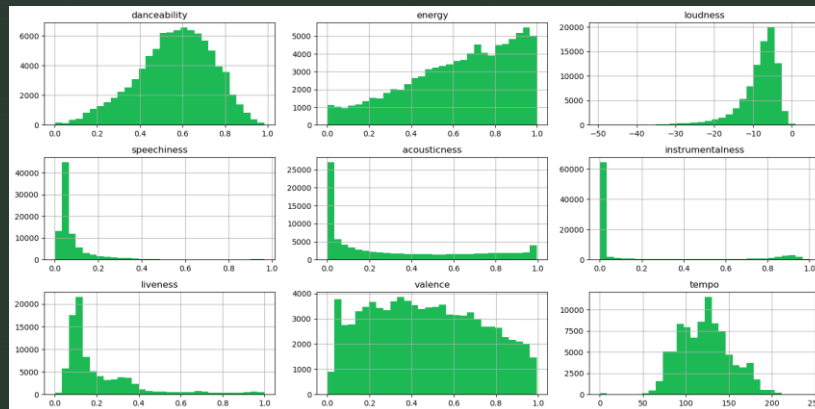
- Las diez canciones más populares del dataset están compuestas, en su mayoría, por canciones de los artistas de mayor popularidad promedio

Popularidad



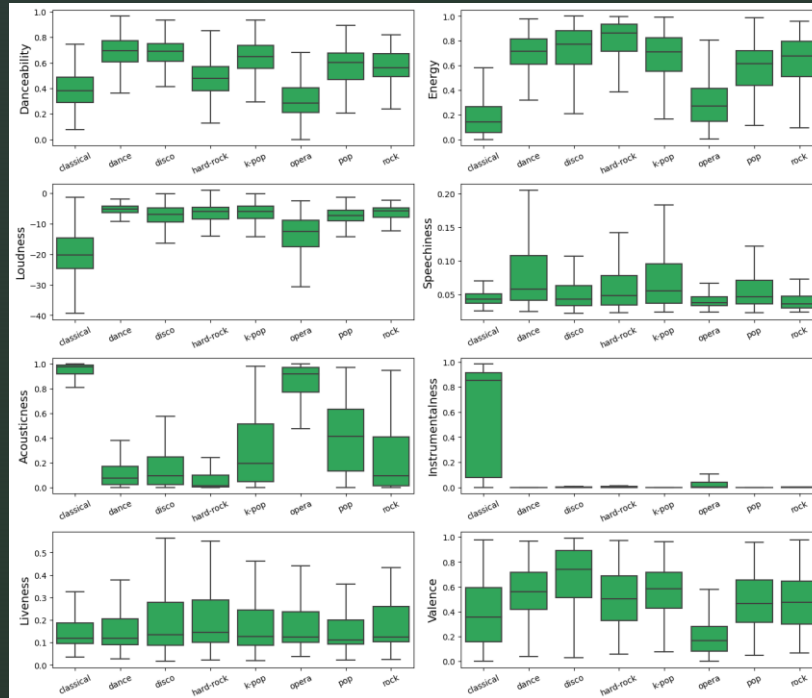
- Si analizamos las canciones más populares y los géneros de dichas canciones, encontramos que en su mayoría pertenecen a los géneros "pop", "dance", "latin" y "latino". Estas canciones, en general son buscadas por su altos valores de "liveness" y "danceability", es decir, son canciones muy bailables.

Histogramas de Frecuencia de Variables



- Histogramas muestran la distribución de cada variable numérica.
- Gran abanico de distribuciones. Muchas variables están alejadas del ideal de distribución normal.
- La estandarización de los datos no es trivial

Comparacion de Variables



- Las variables poseen diferente distribución según el género.
- Géneros relacionados presentan distribuciones similares.
- Géneros disímiles poseen no son comparables en sus variables,
- Estas similitudes y diferencias son útiles para el sist. de recomendación.

Insight

- Las canciones más populares pertenecen, en su gran mayoría, a los géneros pop, hip-hop y latino. En consecuencia, los artistas más populares son los que se dedican a los mencionados géneros.
- Además de los artistas más populares del momento existen artistas de décadas pasadas que permanecen vigentes y escuchados actualmente.
- Las canciones más elegidas por los usuarios son aquellas que transmiten mayor energía, emociones típicamente calificadas como positivas y voluntad de bailar. Sólo el 15 % de las canciones del dataset original son consideradas bailables.
- Casi con exclusividad las piezas musicales poseen voz cantada, sólo un número menor poseen palabra hablada.
- Menos del 9 % de las canciones poseen letras explícitas, no aptas para menores.
- Las variables que caracterizan a las canciones evolucionan con el pasar de los años. En la década del 2010 se detectaron cambios de tendencia en estas variables.

Recomendaciones

- Para atraer a oyentes jóvenes, entre 18 y 30 años, la plataforma debe poseer un amplio repertorio de canciones bailables, que transmitan emociones positivas y energía. Los géneros pop, hip-hop, latinos, entre otros, gozan de popularidad puesto que cumplen con las características mencionadas.
- Se recomienda explorar la posibilidad de ofrecer un servicio diferenciado para menores de edad. Este rango etario posee preferencias musicales similares a los jóvenes de entre 18 y 30 años. Este servicio diferenciado debe poseer la opción de que los padres puedan filtrar las canciones con letras violentas y/o explícitas.
- Por otro lado, a los fines de convencer y/o retener la audiencia de más edad, el servicio Spotify debe ofrecer canciones de artistas consagrados que permanecen vigentes. Artistas tales como Metallica, Michael Jackson, The Smith, 2Pac, Eminem, etcétera, son de géneros dispares y de diferentes décadas pero que figuran en los tops de popularidad.
- Géneros musicales que tuvieron su pico de popularidad en décadas pasadas pero que siguen siendo muy escuchados podrían beneficiarse del influjo de artistas nuevos que siguen produciendo música en estos géneros. Spotify podría dar más visibilidad a dichos artistas para ofrecer novedades para atraer a posibles clientes que gusten de dicha música.
- Tomar en consideración lo más escuchado al momento de modelar el sist. de clasificación y recomendación.
- Analizar diferentes criterios de normalización de los datos.
- Analizar diferentes criterios de normalización de los datos.
- Para el sistema de recomendación: evaluar algoritmos de clustering, como por ejemplo: KMeans, DBSCAN, AGNES, etc.