

Actividad Evaluable: Mapas de calor y boxplots

el conjunto de datos usados en la actividad provienen de Spotify, específicamente son los diferentes valores que se le asignan a cada canción que existe dentro de las bases de datos de Spotify. Este mismo se obtuvo de el siguiente link

<https://www.kaggle.com/bricevergnou/spotify-recommendation>

Este dataset fue creado por un tercero usando el API de Spotify, por lo que las canciones existentes muestran la opinión del creador, y representa 100 canciones que le gustan y 95 que no.

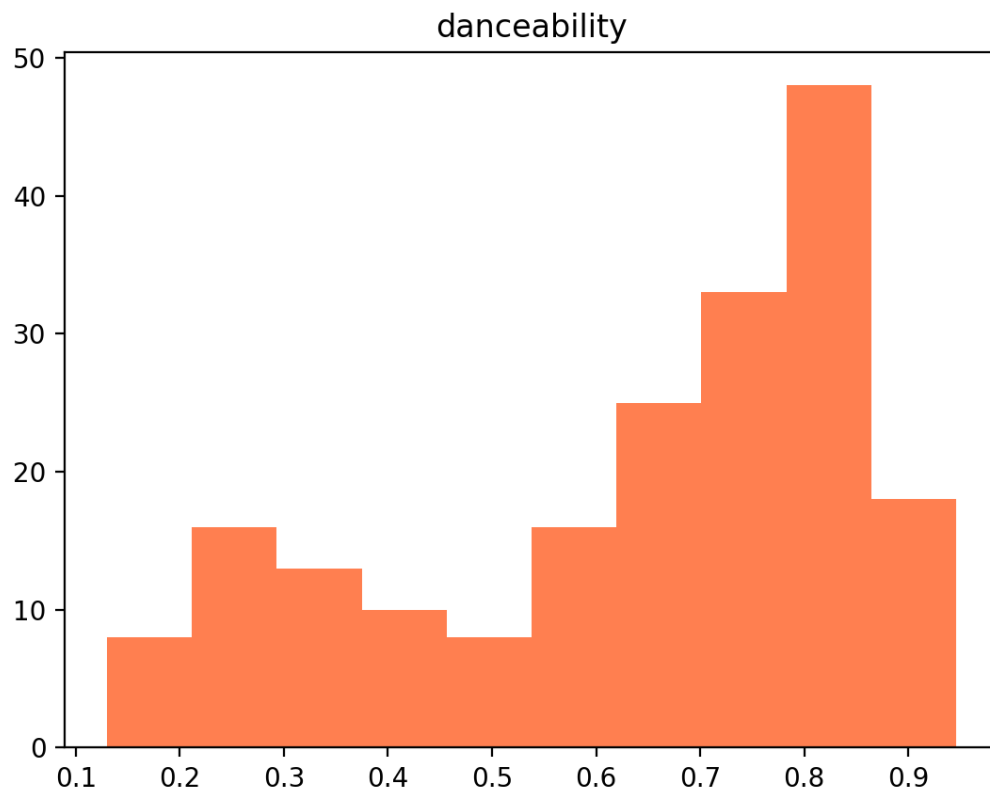
En el dataset existen 195 datos, y 14 variables.

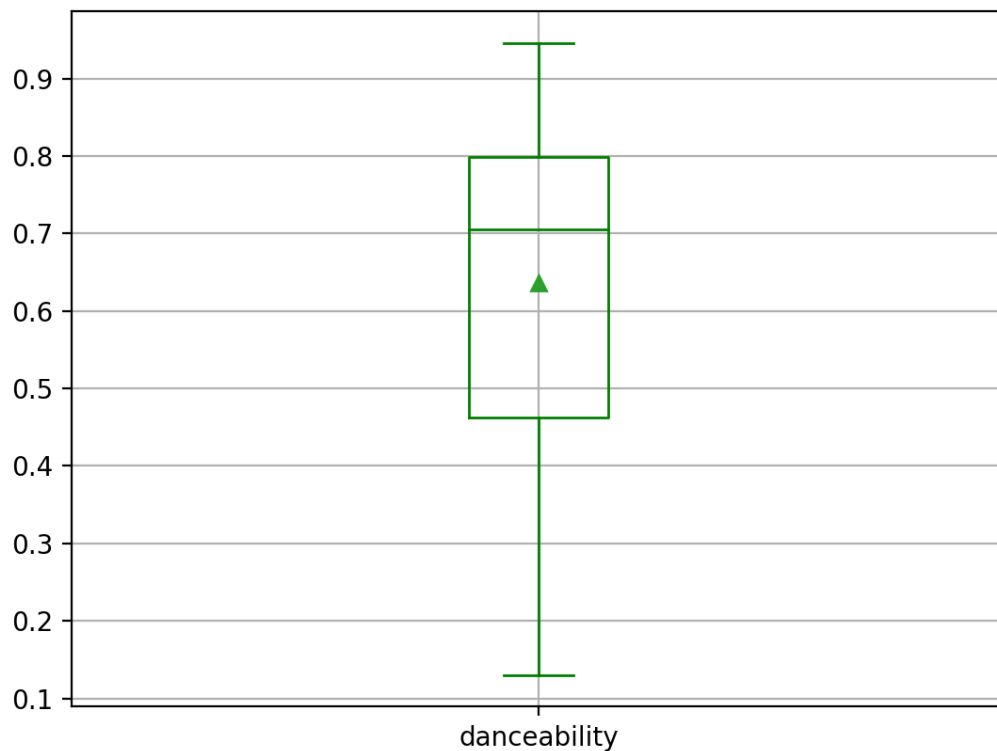
Estas variables son de los siguientes tipos:

danceability	float64
energy	float64
key	int64
loudness	float64
mode	int64
speechiness	float64
acousticness	float64
instrumentalness	float64
liveness	float64
valence	float64
tempo	float64
duration_ms	int64
time_signature	int64
liked	int64

En esta actividad se decidió utilizar las variables “danceability” y “acousticness” ya que son dos partes de la música muy relevantes al determinar si al creador de la lista le interesan las canciones y son partes importantes de toda la música en general.

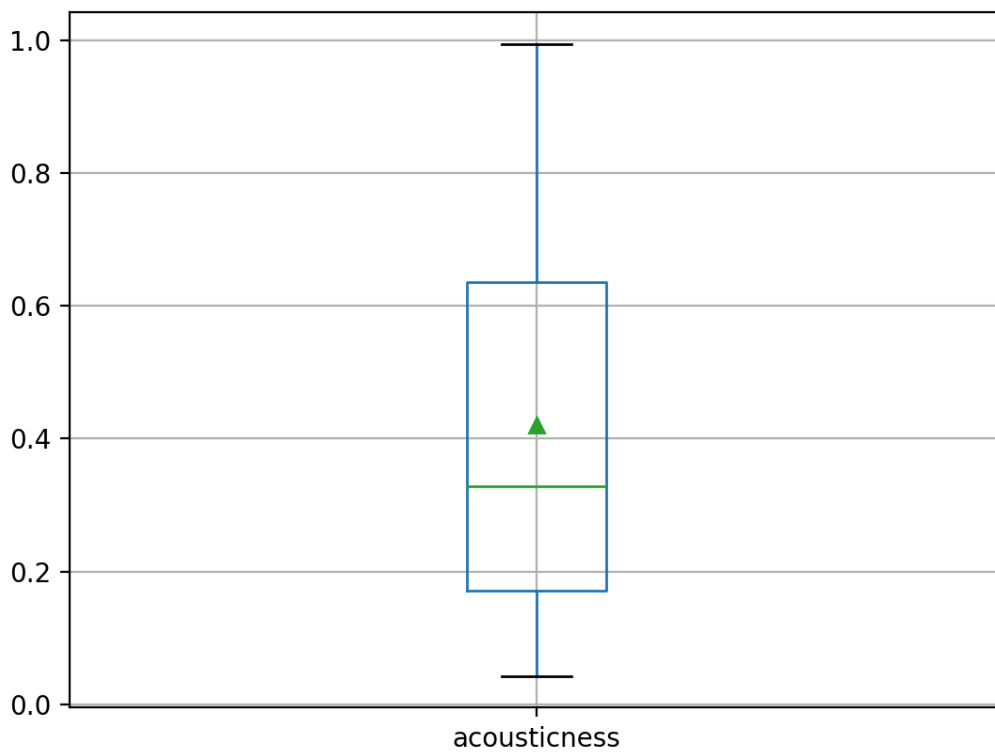
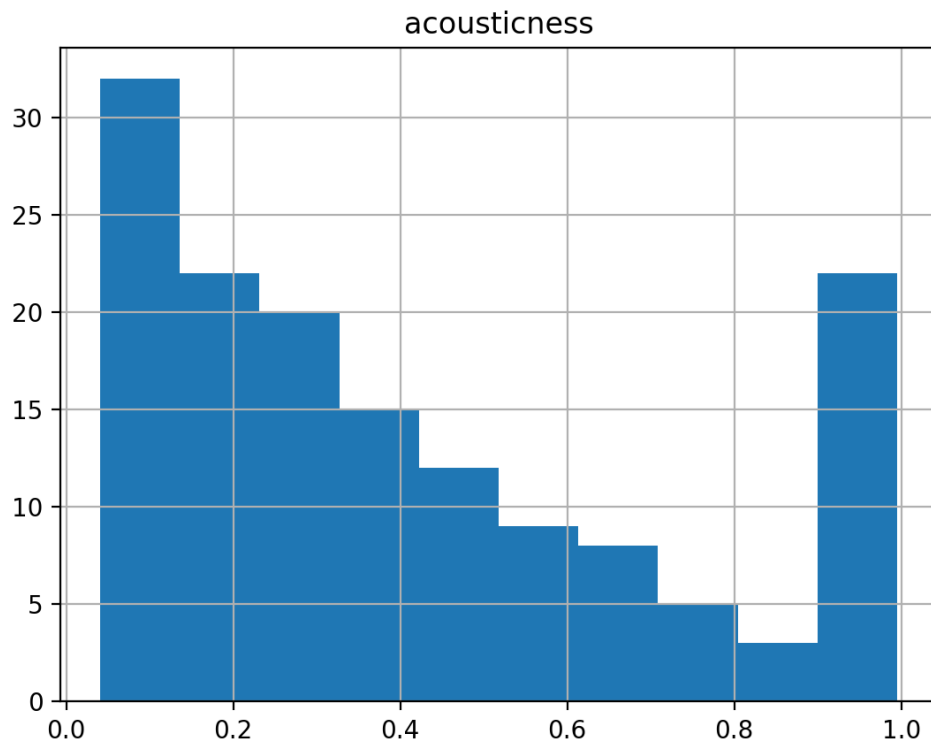
Danceability





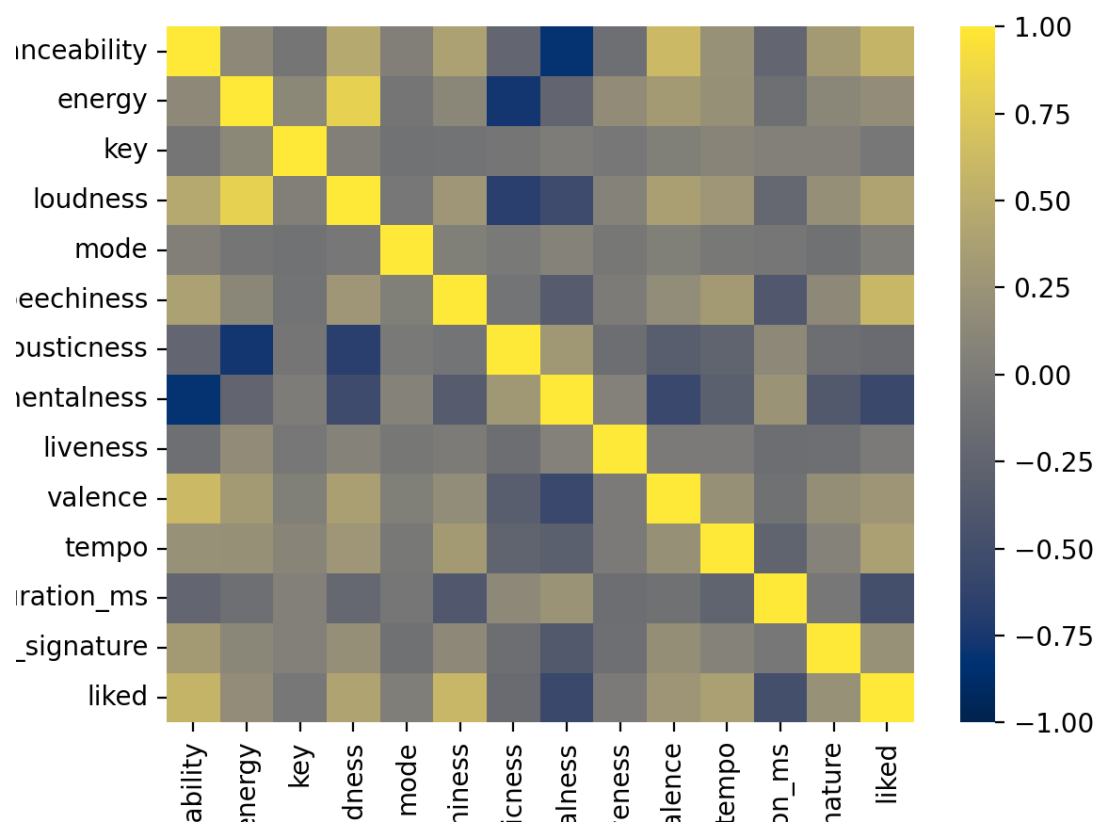
las gráficas anteriores muestran el histograma y el diagrama de bigotes de la variable elegida, en este caso 'danceability', en esta variable existe una variación ligera lo cual es interesante especialmente notando que la mitad de las canciones son aquellas que el creador del dataset no le parecen atractivas. Esto sugiere que 'danceability' no afecta mucho la decisión del creador sobre su interés en la canción específica.

Acousticness



las gráficas anteriores muestran el histograma y el diagrama de bigotes de la variable elegida, en este caso 'danceability', esta variable a diferencia de la anterior muestra datos interesantes especialmente en el histograma, ya que dentro de la variación que existe se pueden notar dos diferentes barras con una gran cantidad de datos, tomando en cuenta la cantidad de canciones positivas y negativas que el autor del dataset eligió, esta variable es significativa al decidir si una canción le será de agrado al creador del dataset.

Mapa de Calor



Por último, el mapa de calor muestra datos significativos en relación a el interés que el el creador del dataset tiene con cada variable. La última columna que muestra a la variable 'liked' sugiere que a el creador le interesa específicamente canciones que se puedan bailar, o en otras palabras con un índice de 'danceability' alto, pero no tan significativamente alto. Le agradan canciones con un alto grado de 'speechiness', y aquellas que tienen un 'tempo' alto.

Aún más interesante que esto son las variables que no le agradan al creador.

Le agradan canciones con un 'duration_ms' bajo, es decir de corta duración, le agradan canciones con poca instrumentación o 'instrumentalness'.

Las variables anteriormente mencionadas son las que muestran más claramente el gusto que el creador del dataset tiene sobre las canciones.