


 An app error occurred. Try re-running the app, or contact the app's owner for help. Details

Librerías

Llamado Dataset

	Unnamed: 0 int64 0 - 9	track_id object 5SuOikwiRy... 10% 4qPNDBW1i... 10% 8 others 80%	artists object Gen Hoshino 10% Ben Woodw... 10% 8 others 80%	album_name obj... We Sing. We ... 20% Comedy 10% 7 others 70%	track_name object Comedy 10% Ghost - Aco... 10% 8 others 80%	popularity int64 55 - 82	duratic 149610
0	0	5SuOikwiRyPMVo...	Gen Hoshino	Comedy	Comedy	73	
1	1	4qPNDBW1i3p13...	Ben Woodward	Ghost (Acoustic)	Ghost - Acoustic	55	
2	2	1iJBSr7s7jYXzM8...	Ingrid Michaelson...	To Begin Again	To Begin Again	57	
3	3	6lfxq3CG4xtTiEg...	Kina Grannis	Crazy Rich Asian...	Can't Help Falling...	71	
4	4	5vjLSffimilP26QG...	Chord Overstreet	Hold On	Hold On	82	
5	5	01MVOI9KtVTNfF...	Tyrone Wells	Days I Will Reme...	Days I Will Reme...	58	
6	6	6Vc5wAMmXdKIA...	A Great Big World...	Is There Anybody...	Say Something	74	
7	7	1EzrEOXmMH3G...	Jason Mraz	We Sing. We Dan...	I'm Yours	80	
8	8	0lktbUcnAGrvDO...	Jason Mraz;Colbi...	We Sing. We Dan...	Lucky	74	
9	9	7k9GuJYLp2Azq...	Ross Copperman	Hunger	Hunger	56	

10 rows, 22 cols 10 / page << < Page 1 of 1 > >> 

Escalado de datos

Este escalado de variables es para poder trabajar con los modelos en un futuro


duration_ms Scaler

	duration_ms int64	duration_ms_sca...
0	230666	0.5131086871
1	149610	0.4456027596
2	210826	0.4990851636
3	201933	0.4923652369
4	198853	0.4899686552

5 rows, 2 cols 10 / page << < Page 1 of 1 > >> 

population Scaler

	popularity int64	popularity_scaled f.
0	73	0.73
1	55	0.55
2	57	0.57
3	71	0.71
4	82	0.82

5 rows, 2 cols 10 / page << < Page 1 of 1 > >> 

Group feature Scaler

Características escaladas: ['danceability', 'energy', 'loudness', 'speechiness', 'acousticness', 'instrumentalness', 'liveness']

Creación de nuevas característica

duration_ms a tiempo_ms_seconds

La duración_ms se calculo para que los valores de milisegundos se pasen a minutos y segundos

```
0    3 minutos con 50 segundos
1    2 minutos con 29 segundos
2    3 minutos con 30 segundos
3    3 minutos con 21 segundos
4    3 minutos con 18 segundos
5    3 minutos con 34 segundos
6    3 minutos con 49 segundos
7    4 minutos con 2 segundos
8    3 minutos con 9 segundos
9    3 minutos con 25 segundos
Name: duration_ms_scaled, dtype: object
```

track_genre a Label Encoding

Cambiar la etiqueta categórica de track_genre a un valor numérico.

```
track_genre  track_genre_encoded
0    acoustic                0
1    acoustic                0
2    acoustic                0
3    acoustic                0
4    acoustic                0
```

Nueva variable "Intensity"

Al tener una buena correlación gracias a la matriz, decidimos optar por juntar la energy con la danceability

```
0    0.316382
1    0.070782
2    0.159637
3    0.016095
4    0.277943
5    0.335968
6    0.060740
7    0.316885
8    0.262690
9    0.283598
Name: intensity, dtype: float64
```

Eliminación de característica

Se elimino.

Unnamed : 0 - artist - album_name - track_id - track_name - explicit - time_signature - mode - duration_ms - log_duration_ms

	popularity int64 55 - 82	danceability float64 0.27005076142131...	energy float64 0.0596 - 0.632	key int64 0 - 11	loudness float64 -5.7370105247581...	speechiness float64 0.0305699481865...	acousticness float64 0.0323 - 0.967
0	73	0.6862944162	0.461	1	0.7913915247	0.1481865285	0.0323
1	55	0.4263959391	0.166	1	0.5973771341	0.07906735751	0.967
2	57	0.4446700508	0.359	0	0.7361226717	0.05772020725	0.213
3	71	0.2700507614	0.0596	0	0.5737010525	0.03761658031	0.967
4	82	0.6274111675	0.443	2	0.7371030095	0.05450777202	0.413
5	58	0.6984771574	0.481	6	0.7532693339	0.1088082902	0.213
6	74	0.4131979695	0.147	2	0.7529918798	0.03678756477	0.813
7	80	0.7137055838	0.444	11	0.743576938	0.04321243523	0.513
8	74	0.6345177665	0.414	0	0.7552485064	0.03823834197	0.213
9	56	0.4487309645	0.632	1	0.7909475982	0.03056994819	0.413

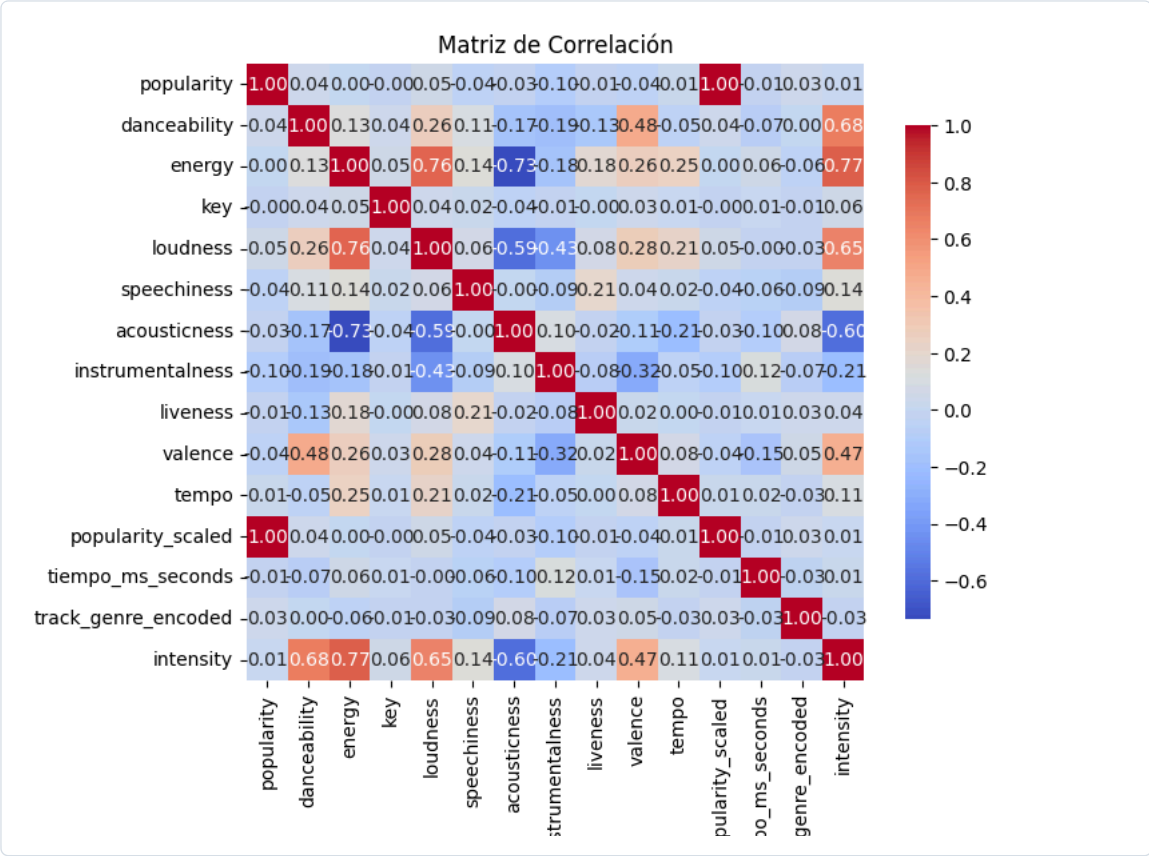
10 rows, 17 cols

10 / page

<< < Page 1 of 1 > >>

↓

Correlation Matrix



Exploración nuevo DataSet "spotify"

Información del DataSet

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 113999 entries, 0 to 113998
Data columns (total 17 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   popularity             113999 non-null  int64
1   danceability           113999 non-null  float64
2   energy                 113999 non-null  float64
3   key                    113999 non-null  int64
4   loudness               113999 non-null  float64
5   speechiness            113999 non-null  float64
6   acousticness           113999 non-null  float64
7   instrumentalness        113999 non-null  float64
8   liveness                113999 non-null  float64
9   valence                113999 non-null  float64
10  tempo                  113999 non-null  float64
11  track_genre            113999 non-null  object
12  duration_ms_scaled     113999 non-null  object
13  popularity_scaled      113999 non-null  float64
14  tiempo_ms_seconds      113999 non-null  float64
15  track_genre_encoded    113999 non-null  int64
16  intensity              113999 non-null  float64
dtypes: float64(12), int64(3), object(2)
memory usage: 14.8+ MB
```

Cantidad total de datos

```
float64    12
int64       3
object      2
Name: count, dtype: int64
```

Futuro

El objetivo de este proyecto es desarrollar un sistema de recomendación de música. Específicamente usando K-Means y K-Vecinos más cercanos (K-NN). Hemos utilizado el procedimiento previo para preparar el conjunto de datos.

- **Escalado de variables:** La mayoría de las variables numéricas han sido escaladas. Para asegurar que todas las características sean similares.
- **Eliminación de variables categóricas:** Las variables categóricas se han eliminado o transformado con el método de Label Encoding, para evitar problemas con los algoritmos que no puedan procesar ese tipo de dato.

El sistema de recomendación funcionará agrupando canciones similares (K-Means) o sugiriendo canciones basadas a las características (K-NN). De esta forma se busca que el sistema pueda recomendar canciones que se alineen con el gusto del usuario.

Cabe recalcar que el método para lograr el objetivo puede cambiar al avanzar con el desarrollo y evaluación de los modelos. Nuestra intención es buscar el mejor modelo para nuestro Recomendador.