Practica de laboratorio Unidad 2:

Introduccion a la limpieza de datos

Materia:

- Intorduccion a la ciencia de datos

Unidad 2:

Procesamiento y Limpieza de Datos

Practica:

Limpieza de una Base de Datos Ensuciada

Alumnos:

- Medina Sánchez Sugey
- Hernández Muñoz Karol

CARGA DE BASE DE DATOS

	id	name	genre	artists	album	popularity	duration_ms	explicit
0	NaN	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	False
1	NaN	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	False
2	NaN	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	False
3	NaN	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	False
4	NaN	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	False

ANALISIS DE BASE DE DATOS:

- Hay 7303 filas y 8 columnas
- La columna 'id' todos los datos son nulos
- La mayoria de las columnas son tipo 'Object'
- Hay 673 filas duplicadas
- Todas las columnas tienen un valor 'invalid' excepto la columna ´'popularity'
- Todas las columnas tienen valores nulos (Nan)

TIPOS DE DATOS DE CADA COLUMNA

Nombre de canciones

name

- Tipo 'object' str texto

Genero de las canciones:

genre

- Tipo ´óbject´ str texto

Artistas

artists

- Tipo 'object' str texto

Album

album

- Tipo 'object' str texto

Popularidad

popularity

- Tipo 'float' decimal

Duracion

duration_ms -Tipo 'float' decimal

Explicito

explicit

- Tipo 'object' str, los datos son categorias:
  - True
  - False
- Informacion clara y facil de comprender

DOCUMENTACION Y REPÓRTE

° ANALISI INICIAL:

```
df.describe()
✓ 0.0s
```

	id	popularity
count	0.0	7089.000000
mean	NaN	30.765411
std	NaN	19.895505
min	NaN	0.000000
25%	NaN	16.000000
50%	NaN	29.000000
75%	NaN	45.000000
max	NaN	90.000000

Resumen estadístico de las columnas numéricas del DataFrame. Esto proporciona información clave como el conteo, la media, la desviación estándar, los valores mínimos y máximos, y los cuartiles 25%, 50%, y 75%. Esto es útil para entender la distribución y las características básicas de los datos, lo que puede ayudar en el análisis y la toma de decisiones en un informe.

```
#Cantidad de valores nulos por columna
df.isnull().sum()
```
✓ 0.0s

```
id              7303
name              81
genre             81
artists           63
album             71
popularity       214
duration_ms       62
explicit          79
dtype: int64
```

Valores faltantes:

Permite identificar la cantidad total de valores faltantes en cada columna del DataFrame. Al aplicar esta función, se obtiene un resumen que muestra cuántos datos están ausentes, lo que es fundamental para, evaluar la calidad de los datos antes de realizar análisis más profundos.

°- Como podemos apreciar ID cuenta con todos los valores nulos en el Dataframe.

TOTAL DE FILAS DUPLICADAS ENCONTRADAS:

Es una función esencial para la limpieza de datos, ya que proporciona una evaluación clara de la cantidad de registros duplicados, lo que permite tomar decisiones informadas sobre cómo mejorar la calidad y la utilidad del dataset.

°- Como pudimos apreciar contamos con 673 duplicados en el dataframe.

```
df.duplicated().sum()
```
✓ 0.0s

```
np.int64(673)
```

DESCRIPCION DE LOS TIPOS DE DATOS ORIGINALES Y LOS PROBLEMAS ENCONTRADOS:

```
#Informacion del data frame
df.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7303 entries, 0 to 7302
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype  
---  ------       --------------  -----  
 0   id           0 non-null      float64
 1   name         7222 non-null   object 
 2   genre        7222 non-null   object 
 3   artists      7240 non-null   object 
 4   album        7232 non-null   object 
 5   popularity   7089 non-null   float64
 6   duration_ms  7241 non-null   object 
 7   explicit     7224 non-null   object 
dtypes: float64(2), object(6)
memory usage: 456.6+ KB
```

- Es una función esencial para la exploración inicial de un DataFrame, ya que brinda un resumen útil para comprender la estructura de los datos y planificar los pasos siguientes en el análisis.
- Pudimos observar que la columna de duración tiene un problema, ya que esta en tipo objeto y debe ser numero real.

PROCESO DE LIMPIEZA:

• Eliminación de duplicados

° Con los duplicados primero rectificamos la cantidad existente en el dataframe.

° Con el siguiente código (df.duplicates()) hicimos la eliminación de duplicados en el dataframe, poniéndolos en el dataframe 2.

Finalmente rectificamos que ya no estén valores duplicados en el dataframe que hayamos agregado.

Eliminacion de duplicados

```
df.duplicated().sum()
```

✓ 0.0s

```
np.int64(673)
```

```
#Eliminar duplicados
df2=df.drop_duplicates()
df2
```

✓ 0.0s

	id	name	genre	artists	album	popularity	duration_ms	explicit
0	NaN	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	False
1	NaN	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	False
2	NaN	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	False
3	NaN	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	False
4	NaN	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	False
...	...	...	...	...	...	...	...	...
7296	NaN	invalid	disco	Cameron Vigano	Discovering The World	44.0	100128.0	False
7298	NaN	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	False
7299	NaN	invalid	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	False
7300	NaN	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	invalid	38.0	237013.0	False
7301	NaN	PART OF ME HARDSTYLE	invalid	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	False

6630 rows × 8 columns

```
df2.duplicated().sum()
```

✓ 0.0s

```
np.int64(0)
```

ELIMINACION DE COLUMNA:

Eliminacion de columnas

```
#Eliminar una columna
df2=df2.drop(columns=['id'])
df2
```

95] ✓ 0.0s

..

	name	genre	artists	album	popularity	duration_ms	explicit
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	False
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	False
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	False
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	False
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	False
...	...	...	...	...	...	...	...
7296	invalid	disco	Cameron Vigano	Discovering The World	44.0	100128.0	False
7298	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	False
7299	invalid	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	False
7300	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	invalid	38.0	237013.0	False
7301	PART OF ME HARDSTYLE	invalid	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	False

6630 rows × 7 columns

- este proceso checamos anteriormente que toda la columna son valores nulos, entonces con el código que vemos en pantalla nos ayuda a eliminar dicha columna que nos favorece en nuestro análisis.

RENOMBRACION DE COLUMNAS

```
#Renombrar columnas
df3 = df2.rename(columns={'name': 'Nombre',
                          'genre':'Genero',
                          'artists':'Artistas',
                          'album':'Album',
                          'popularity':'Popularidad',
                          'duration_ms':'Duracion',
                          'explicit':'Explicito'})
df3
```

✓ 0.0s

	Nombre	Genero	Artistas	Album	Popularidad	Duracion	Explicito
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	False
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	False
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	False
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	False
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	False
...	...	...	...	...	...	...	...
7296	invalid	disco	Cameron Vigano	Discovering The World	44.0	100128.0	False
7298	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	False
7299	invalid	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	False
7300	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	invalid	38.0	237013.0	False
7301	PART OF ME HARDSTYLE	invalid	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	False

6630 rows × 7 columns

```
#Traducir 'false' y 'true'
Tradexplicito= {
    'True':'Verdadero',
    'False':'Falso'
}
```

✓ 0.0s

En este paso traducimos las columnas del dataframe 2 a un nuevo dataframe (df3).
Igualmente traducimos el contenido dentro de la columna (explicito) ya que es la única columna que se pudiera modificar al español

```
#Traducimos las filas de la columna 'Explicito'
df3['Explicito']=df3['Explicito'].replace(Tradexplicito)
df3
```

✓ 0.0s

	Nombre	Genero	Artistas	Album	Popularidad	Duracion	Explicito
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	Falso
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	Falso
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	Falso
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	Falso
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	Falso
...	...	...	...	...	...	...	...
7296	invalid	disco	Cameron Vigano	Discovering The World	44.0	100128.0	Falso
7298	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	Falso
7299	invalid	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	Falso
7300	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	invalid	38.0	237013.0	Falso
7301	PART OF ME HARDSTYLE	invalid	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	Falso

6630 rows × 7 columns

CORRECION DE TIPOS DE DATOS

```
#Informacion del data frame 3
df3.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
Index: 6630 entries, 0 to 7301
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Nombre       6549 non-null   object
 1   Genero       6549 non-null   object
 2   Artistas     6567 non-null   object
 3   Album        6559 non-null   object
 4   Popularidad  6416 non-null   float64
 5   Duracion     6568 non-null   object
 6   Explicito    6551 non-null   object
dtypes: float64(1), object(6)
memory usage: 414.4+ KB
```

```
#Cambiamos la columna 'Duracion' a Decimal
df3['Duracion']=pd.to_numeric(df3['Duracion'], errors='coerce')
df3.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
Index: 6630 entries, 0 to 7301
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Nombre       6549 non-null   object
 1   Genero       6549 non-null   object
 2   Artistas     6567 non-null   object
 3   Album        6559 non-null   object
 4   Popularidad  6416 non-null   float64
 5   Duracion     6415 non-null   float64
 6   Explicito    6551 non-null   object
dtypes: float64(2), object(5)
```

Con el análisis previo, descubrimos que solo hay una columna que necesita este código para la corrección del tipo de datos que tiene (duración) que esta en tipo ‘object’ (objecto), lo pasamos a ‘float’ (decimal-real).

GRAFICOS DE BLOXPLOT

```
#Creamos un df alternativo para poder crear el boxplot
df0=df3.dropna()
df0
```

✓ 0.0s

	Nombre	Genero	Artistas	Album	Popularidad	Duracion	Explicito
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	Falso
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	Falso
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	Falso
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	Falso
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	Falso
...	...	...	...	...	...	...	...
7296	invalid	disco	Cameron Vigano	Discovering The World	44.0	100128.0	Falso
7298	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	Falso
7299	invalid	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	Falso
7300	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	invalid	38.0	237013.0	Falso
7301	PART OF ME HARDSTYLE	invalid	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	Falso

5862 rows × 7 columns

En esta parte creamos un dataframe alternativo (df0) para poder crear el boxplot , además de que eliminamos los valores nulos temporalmente para realizarlo. Hay que si no los eliminamos tal no agarraría el grafico.

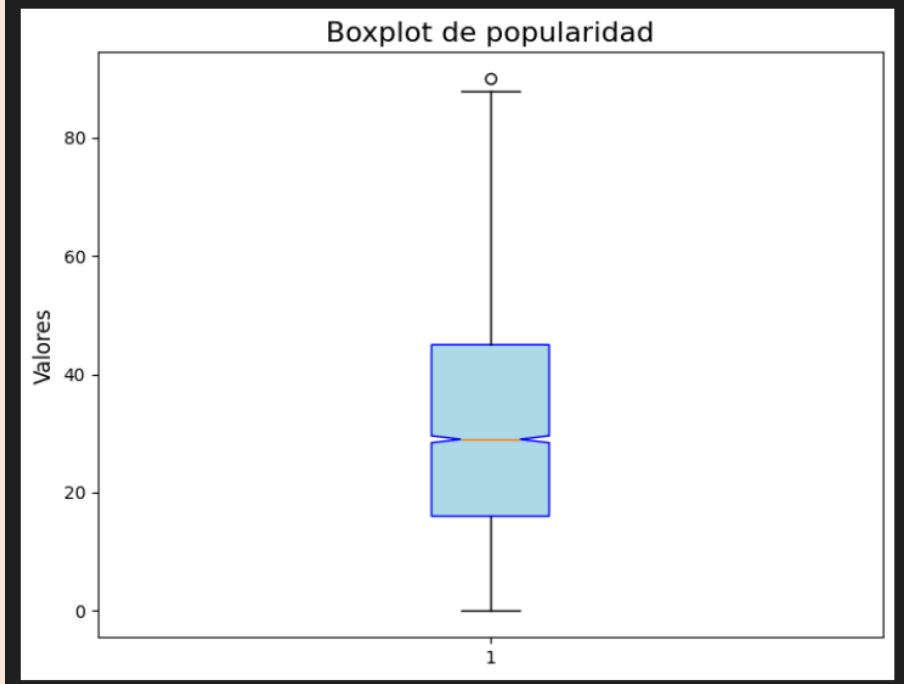
GRAFIQUITOSSS

```python
#Crear boxplot
plt.figure(figsize=(8, 6))
plt.boxplot(df0['Popularidad'],
            patch_artist=True,
            notch=True,
            boxprops=dict(facecolor='lightblue', color='blue'))

# Personalizar el gráfico
plt.title('Boxplot de popularidad', fontsize=16)
plt.ylabel('Valores', fontsize=12)

# Mostrar el gráfico
plt.show()
```

✓ 0.5s

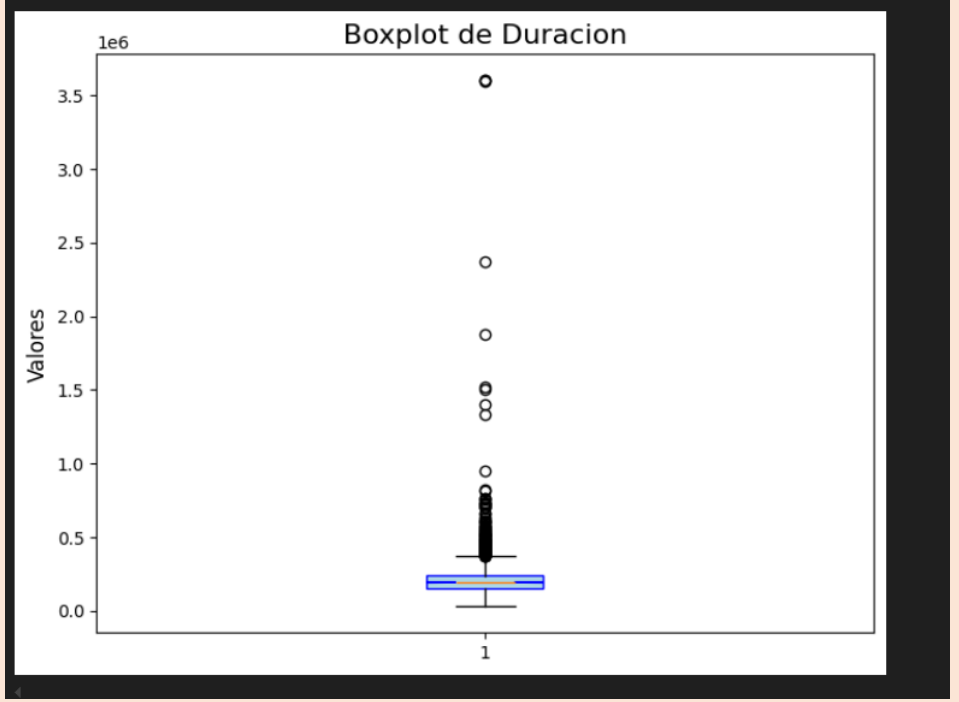



```python
#Crear boxplot
plt.figure(figsize=(8, 6))
plt.boxplot(df0['Duracion'],
            patch_artist=True,
            notch=True,
            boxprops=dict(facecolor='lightblue', color='blue'))

# Personalizar el gráfico
plt.title('Boxplot de Duracion', fontsize=16)
plt.ylabel('Valores', fontsize=12)

# Mostrar el gráfico
plt.show()
```

✓ 0.1s



SUSTITUCION Y ELIMINACION DE LOS NAN

```
#Codigo para reemplazar los 'invalid' por 'Nan'
df3.replace('invalid',np.nan,inplace=True)
```
✓ 0.0s

- Con este pequeño código, sustituimos los ‘invalid’ de nuestro dataframe, por valores nulos.

```
#Cantidad de valores nulos por columna despues de cambiar los invalid a nulos
df3.isnull().sum()
```
✓ 0.0s

```
Nombre        209
Genero        227
Artistas      215
Album         216
Popularidad   214
Duracion      215
Explicito     236
dtype: int64
```

Como pudimos apreciar todos los datos ‘invalid’ fueron cambiados por valores nulos.

```
#Cambiar los valores nulos (Nan) por:

#De la colmuna "Genero" por "Sin_genero"
df3['Genero'].fillna("Sin genero",inplace=True)
#De la columna "Nombre" por "Sin_nombre"
df3['Nombre'].fillna("sin_nombre",inplace=True)

#De la colmuna "Artistas" por "Sin_artistas"
df3['Artistas'].fillna("sin_artista",inplace=True)

#De la colmuna "Album" por "Sin_album"
df3['Album'].fillna("sin_album",inplace=True)

##De la colmuna "Explicito" por "sin_valor"
df3['Explicito'].fillna("sin_valor",inplace=True)

df3
```
✓ 0.0s

Con este código simplemente cambiamos los valores nulos de las columnas que son tipo carácter por un texto que nos sea conveniente en el mismo data frame 3.

```
#Cantidad de valores invalidos luego de  sustituir
df3.isnull().sum()
```

✓ 0.0s

```
Nombre          0
Genero          0
Artistas        0
Album           0
Popularidad    214
Duracion       215
Explicito        0
dtype: int64
```

```
#Eliminamos filas de las columnas 'Genero'
df3=df3.dropna()
df3
```

✓ 0.0s

	Nombre	Genero	Artistas	Album	Popularidad	Duracion	Explicito
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	Falso
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	Falso
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	Falso
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	Falso
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	Falso
...	...	...	...	...	...	...	...
7296	sin_nombre	disco	Cameron Vigano	Discovering The World	44.0	100128.0	Falso
7298	Happy Day In Hell - Hazbin Hotel Original Soun...	happy	Erika Henningsen, Stephanie Beatriz, Sam Haft,...	Happy Day In Hell (Hazbin Hotel Original Sound...	58.0	177664.0	Falso
7299	sin_nombre	garage	King Gizzard & The Lizard Wizard	12 Bar Bruise	18.0	149373.0	Falso
7300	Hard Rock Bottom of Your Heart	hard-rock	Randy Travis	sin_album	38.0	237013.0	Falso
7301	PART OF ME HARDSTYLE	Sin genero	SICK LEGEND	PART OF ME HARDSTYLE	58.0	131657.0	Falso

6208 rows × 7 columns

- En el primer código checamos que se hayan sustituido correctamente los valores nulos.
- En el segundo código eliminamos los valores nulos de las columnas que son tipo decimales.

```
#Cantidad de datos nulos luego de limpiarlos y sustituir los importantes
df3.isnull().sum()
```

✓ 0.0s

```
Nombre          0
Genero          0
Artistas        0
Album           0
Popularidad     0
Duracion        0
Explicito       0
dtype: int64
```

Finalmente podemos observar que ya no existen datos nulos en el dataframe.

ELIMINACION DE DATOS ATIPICOS

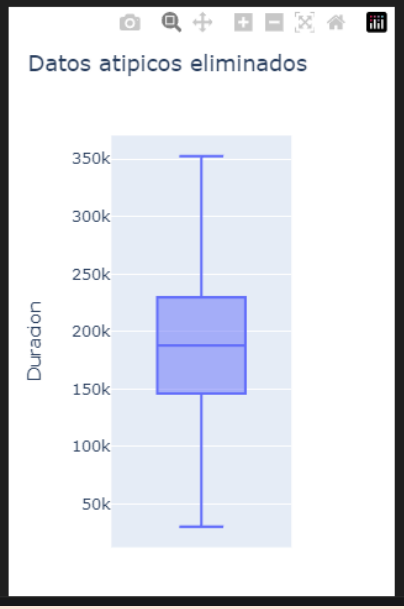
```
#Definimos cuartil 75 y restamos el cuartil 75
iqr= df3['Duracion'].quantile(0.75)-df3['Duracion'].quantile(0.25)

#Desarrollamos los filtros superior o inferior
filtro_inferior= df3['Duracion']>df3['Duracion'].quantile(0.25)-(iqr* 1.5)
filtro_superior= df3['Duracion']<df3['Duracion'].quantile(0.75)+(iqr* 1.3)

df3_filtrado= df3[filtro_inferior & filtro_superior]

#Graficando el boxplot
fig= px.box(df3_filtrado, y='Duracion', title='Datos atipicos eliminados')
fig.update_layout(width=300,height=500)
fig.show()
```

0.0s



De los boxplot anteriores, solo es necesario eliminar los datos atípicos de la columna ‘DURACION’, y podemos apreciar una grafica con datos atípicos completamente eliminados.

RESULTADOS:

Base de datos completamente limpiecitaaaaaa!!!!

```
#Base de datos finalizada
df3.head(5)
```

✓ 0.0s

	Nombre	Genero	Artistas	Album	Popularidad	Duracion	Explicito
0	Acoustic	acoustic	Billy Raffoul	1975	58.0	172199.0	Falso
1	Acoustic	acoustic	Billy Raffoul	A Few More Hours at YYZ	57.0	172202.0	Falso
2	Here Comes the Sun - Acoustic	acoustic	Molly Hocking, Bailey Rushlow	Here Comes the Sun (Acoustic)	42.0	144786.0	Falso
3	Acoustic #3	acoustic	The Goo Goo Dolls	Dizzy up the Girl	46.0	116573.0	Falso
4	My Love Mine All Mine - Acoustic Instrumental	acoustic	Guus Dielissen, Casper Esmann	My Love Mine All Mine (Acoustic Instrumental)	33.0	133922.0	Falso

```
df3.shape
```

✓ 0.0s

```
(6208, 7)
```

```
df3.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
Index: 6208 entries, 0 to 7301
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Nombre       6208 non-null   object
 1   Genero       6208 non-null   object
 2   Artistas     6208 non-null   object
 3   Album        6208 non-null   object
 4   Popularidad  6208 non-null   float64
 5   Duracion     6208 non-null   float64
 6   Explicito    6208 non-null   object
dtypes: float64(2), object(5)
memory usage: 388.0+ KB
```

Comprobación de base de datos limpia.
Guardamos la base limpia en archivo csv.

```
df3.to_csv("limpiecitaaaaaaa.csv")
```

✓ 0.0s

```
df3.info()
```

✓ 0.0s

```
<class 'pandas.core.frame.DataFrame'>
Index: 6208 entries, 0 to 7301
Data columns (total 7 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   Nombre       6208 non-null   object
 1   Genero       6208 non-null   object
 2   Artistas     6208 non-null   object
 3   Album        6208 non-null   object
 4   Popularidad  6208 non-null   float64
 5   Duracion     6208 non-null   float64
 6   Explicito    6208 non-null   object
dtypes: float64(2), object(5)
memory usage: 388.0+ KB
```

```
df3.isnull().sum()
```

✓ 0.0s

```
Nombre         0
Genero         0
Artistas       0
Album          0
Popularidad    0
Duracion       0
Explicito      0
dtype: int64
```

```
df3.duplicated().sum()
```

✓ 0.0s

```
np.int64(0)
```