

Anexo III. Obtención del artista principal de las canciones

A continuación se presenta el código empleado para obtener el artista principal de todas las canciones de nuestro conjunto de datos.

Para un posible análisis o agrupación en función de los artistas, es necesario obtener información clara y limpia. Esto no es posible con el conjunto de datos actual, ya que la variable 'artists' es una cadena de texto que separa los múltiples artistas que puede tener una canción por comas. En consecuencia, vamos a reducir esta información quedándonos con el artista principal.

Se ha adoptado el criterio de asumir que el primer artista nombrado es el principal, ya que no se aprecia ningún orden alfabético u otro criterio de ordenación alternativo.

```
import pandas as pd
df = pd.read_csv('outputpop.csv', sep=',', encoding = 'ISO-8859-1')
df.head()

c = 0
for valor in df['artists']:
    valor = re.split(r',',valor)
    valor = valor[0]
    valor = re.findall(r'\s*\w+\s*',valor)
    res = ''
    for v in valor:
        res += v
    valor = res
    df['artists'][c] = valor
    c += 1
    if c== 100 or c == 1000 or c == 100000 or c== 250000 or c == 500000 or c ==
750000 or c== 1000000 or c == 1250000:
        print(c)
df.head()

df.to_csv('datospopart1.csv', sep = ',')
```