

# Nubes\_de\_palabras

May 23, 2023

## 1 Nubes de palabras con Python

Para generar la nube de palabras usamos la librería [wordcloud](#).

En los ejemplos que mostramos también usamos PIL (ya vista en la práctica) y NumPy.

### 1.1 ¿Qué es NumPy?

Es una librería muy utilizada y la base de pandas.

Tiene múltiples usos. Principalmente nos permite trabajar con arreglos multidimensionales.

Acá dejamos un ejemplo simple de manipulación de imágenes on NumPy que puede serles de interés como para introducirse en el uso de esta librería: [artículo en https://realpython.com/numpy-tutorial/](https://realpython.com/numpy-tutorial/)

### 1.2 Creamos nubes de palabras

Vamos a crear unas nubes de palabras con los nombres de los artistas del reporte obtenido de Spotify: <https://spotifycharts.com/regional/ar/weekly/latest>

#### 1.2.1 Vamos a utilizar la librería wordcloud

Sitio: [https://github.com/amueller/word\\_cloud](https://github.com/amueller/word_cloud)

Primero procesamos el dataset:

```
[1]: import pandas as pd

ds = pd.read_csv('regional-ar-daily-latest.csv', encoding='utf-8')
ds
```

```
[1]:      Unnamed: 0      Unnamed: 1 \
0      Position      Track Name
1           1      Todo De Ti
2           2      Miénteme
3           3      Qué Más Pues?
4           4      Pareja Del Año
..      ...      ...
196      196      Barrio Prendido
197      197      3G (feat. Jon Z, Don Chezina, Chench...
```

198	198	Tu Veneno
199	199	Adiós
200	200	Si Me Tomo una Cerveza - Remix

```

      Unnamed: 2  \
0           Artist
1   Rauw Alejandro
2           TINI
3           J Balvin
4   Sebastian Yatra
..          ...
196  The La Planta
197           Wisin
198           J Balvin
199  Gustavo Cerati
200           Migrantes

```

Note that these figures are generated using a formula that protects against any artificial inflation of chart positions. \

0	Streams
1	653541
2	437474
3	369430
4	323876
..	...
196	26318
197	26296
198	26230
199	26008
200	25913

```

      Unnamed: 4
0           URL
1   https://open.spotify.com/track/4fSIb4hd0Q151TI...
2   https://open.spotify.com/track/0c0a970mzTWAxKt...
3   https://open.spotify.com/track/6hf0RpxTb0prT5n...
4   https://open.spotify.com/track/2rCb19naJYhaxjL...
..          ...
196 https://open.spotify.com/track/2I3qA0KslRCxFTf...
197 https://open.spotify.com/track/6dyUkkcyYuDiH0w...
198 https://open.spotify.com/track/5KFHrrBMyGj5MhV...
199 https://open.spotify.com/track/1rh4kDY9T4fHVDu...
200 https://open.spotify.com/track/7oySelE0Cb5q1TB...

```

[201 rows x 5 columns]

### 1.2.2 Depuramos el dataset

```
[2]: ds.columns = ds.iloc[0]
```

```
[3]: artistas = ds[1:]
      artistas
```

```
[3]: 0    Position                                Track Name \
1         1                                Todo De Ti
2         2                                Miénteme
3         3                        Qué Más Pues?
4         4                        Pareja Del Año
5         5      YO SE QUE TU (feat. Rusherking)
..      ...
196      196                        Barrio Prendido
197      197  3G (feat. Jon Z, Don Chezina, Chencho Corleone...
198      198                        Tu Veneno
199      199                        Adiós
200      200      Si Me Tomo una Cerveza - Remix
```

```
0          Artist Streams \
1    Rauw Alejandro  653541
2          TINI      437474
3          J Balvin  369430
4    Sebastian Yatra  323876
5          FMK      311363
..      ...
196    The La Planta   26318
197          Wisin    26296
198          J Balvin  26230
199    Gustavo Cerati  26008
200    Migrantes      25913
```

```
0          URL
1  https://open.spotify.com/track/4fSIb4hd0Q151TI...
2  https://open.spotify.com/track/0c0a970mzTWAxKt...
3  https://open.spotify.com/track/6hf0RpxTb0prT5n...
4  https://open.spotify.com/track/2rCbl9naJYhaxjL...
5  https://open.spotify.com/track/4M4DTyLupKvbJKl...
..      ...
196  https://open.spotify.com/track/2I3qAOKslRCxFTf...
197  https://open.spotify.com/track/6dyUkkcyYuDiH0w...
198  https://open.spotify.com/track/5KFHrrBMyGj5MhV...
199  https://open.spotify.com/track/1rh4kDY9T4fHVDu...
200  https://open.spotify.com/track/7oySe1EOCb5q1TB...
```

```
[200 rows x 5 columns]
```

### 1.2.3 Obtenemos los nombres de los artistas

Para esto, obtengamos los valores únicos de la columna “Artist”.

```
[12]: palabras = artistas["Artist"].unique()
```

```
[13]: palabras
```

```
[13]: array(['Rauw Alejandro', 'TINI', 'J Balvin', 'Sebastian Yatra', 'FMK',  
          'KHEA', 'Nio Garcia', 'Los Legendarios', 'Mora', 'Rusherking',  
          'Maria Becerra', 'Duki', 'KAROL G', 'Olivia Rodrigo', 'Bizarrap',  
          'Selena Gomez', 'Myke Towers', 'Crissin', 'L-Gante', 'Rochy RD',  
          'BTS', 'Bruno LC', 'Boza', 'Sech', 'Bad Bunny', 'Big Apple',  
          'Papu DJ', 'Asan', 'Damas Gratis', 'Luck Ra', 'Lil Nas X',  
          'Nicki Nicole', 'Justin Bieber', 'Milly', 'Daddy Yankee',  
          'Natti Natasha', 'Tiago PZK', 'Kali Uchis', 'Dillom', 'Dua Lipa',  
          'Justin Quiles', 'Camilo', 'Migrantes', 'Alex Rose',  
          'Manuel Turizo', 'Nicky Jam', 'The Weeknd', 'Ozuna', 'Maluma',  
          'Doja Cat', 'C. Tangana', 'CNCO', 'Emilia', 'Harry Styles',  
          'La K'onga', 'Bruno Mars', 'Pablo Alborán', 'Cazzu', 'j mena',  
          'Ciro y los Persas', 'Arcangel', 'Wisin', 'Intoxicados', 'Gera MX',  
          'Carlos Vives', 'Juhn', 'Nathy Peluso', 'El Reja', 'Papichamp',  
          'Trueno', 'Luis Alberto Spinetta', 'Airbag', 'Luis Miguel',  
          'Jay Wheeler', 'Jossef', 'Kevin Roldan', 'Andrés Calamaro',  
          'Lyanno', 'Fer Palacio', 'Farruko', 'Billie Eilish',  
          'Lenny Tavárez', 'Gustavo Cerati', 'Miley Cyrus', 'Bella Poarch',  
          'La Joaqui', 'Masked Wolf', 'Ovy On The Drums', 'Soda Stereo',  
          'Piso 21', 'Divididos', 'The Police', 'Reik', 'Los Tipitos',  
          'AURORA', 'Callejeros', 'Patricio Rey y sus Redonditos de Ricota',  
          'Los Abuelos De La Nada', 'Micro TDH', 'Jhay Cortez',  
          'Paulo Londra', 'Jeeiph', 'Duncan Laurence', 'Pappo',  
          'Danny Ocean', 'Turf', 'Tiësto', 'Tones And I', 'Lewis Capaldi',  
          'Locura Mix', 'The La Planta'], dtype=object)
```

## 1.3 Ahora si, armamos la nube de palabras

Ejemplo sacado de <https://towardsdatascience.com/simple-wordcloud-in-python-2ae54a9f58e5>

```
[9]: import numpy as np  
     from PIL import Image
```

```
[10]: from wordcloud import WordCloud, STOPWORDS  
      import matplotlib.pyplot as plt  
  
      # Define a function to plot word cloud  
      def plot_cloud(wordcloud):  
          # Set figure size  
          plt.figure(figsize=(40, 30))
```

```

# Display image
plt.imshow(wordcloud)
# No axis details
plt.axis("off");

text = str(palabras)
#text = str([c.replace(' ','_') for c in palabras])

# Generate wordcloud
wordcloud = WordCloud(width = 300, height = 200, random_state=1,
↳background_color='white', colormap='Set2', collocations=False, stopwords =
↳STOPWORDS).generate(text)
# Plot
plot_cloud(wordcloud)

```



#### 1.4 Nubes con forma

También podemos utilizar una imagen para usar como máscara para “darle forma” a la nube de palabras.

```

[11]: # Import image to np.array
mask = np.array(Image.open('upvote.png'))

```

```
# Generate wordcloud
wordcloud = WordCloud(width = 3000, height = 2000, random_state=1,
    ↳background_color='white', colormap='Set2', collocations=False, stopwords =
    ↳STOPWORDS, mask=mask).generate(text)
# Plot
plot_cloud(wordcloud)
```



Más info en <https://www.datacamp.com/community/tutorials/wordcloud-python>