

Vsevolod Vasylets. Proyecto final del curso:

## "Noticias actuales en un una nube de palabras"

In [2]:

```
#Librerías básicas
import numpy as np
import pandas as pd

#Librerías necesarias para scrapear textos de páginas web
import requests
from bs4 import BeautifulSoup

#Librerías necesarias para abrir imágenes, generar nube de palabras y plot
from PIL import Image
from wordcloud import WordCloud, ImageColorGenerator
import matplotlib.pyplot as plt

#Librerías necesarias para la limpieza de datos
import string
import nltk
from nltk.corpus import stopwords

websites = ['https://www.20minutos.es/', 'https://elpais.com/']
texto = ""

for site in websites:
    result = requests.get(site)
    content = result.text
    soup = BeautifulSoup(content, 'lxml')

    if site == 'https://www.20minutos.es/':
        box = soup.find_all('h1')

    elif site == 'https://elpais.com/':
        box = soup.find_all('h2')

    else:
        print("Error interno. Nunguna web seleccionada")

    for i in box:
        text = i.find('a')
        text = str(text)
        text = text.split('>')[1]
        print("-", text)
        texto += text
```

- Campaña de cartas bomba: interceptan envíos a Defensa y la base de Torrejón y Pedro Sánchez recibió otra la semana pasada</a>
- Interior dice que las cartas bomba son de "fabricación casera" y no descarta que se hayan enviado más</a>
- De color marrón, con la misma letra... Así son las cartas bomba que la Audiencia investiga como posibles delitos de terrorismo</a>
- <strong>Instalaza,</strong> uno de los objetivos de las cartas: fabrica los lanzagr anadas C-90 que España envió a Ucrania </a>

- 30 años, derecho de descarte... en qué consiste el **nuevo plan de Escrivá para las pensiones, a quién beneficia más y a quién perjudica**
- Así es la **subida que aplicará Movistar** a casi todas sus tarifas de sde el día 13 enero
- **¿Una Navidad sin cordero?**: "Con este precio, si cenamos huevos fritos y patatas, no pasa nada"
- 'Pacifiction', Boyero, C. Tangana y otros olvidos imperdonables de los Goya
- **Aitana, sobre sus inicios con Miguel Bernardeau:** "Tras la primera cita, pensé que no nos veríamos nunca más"
- **Pablo Iglesias, contra Pablo Motos:** "Él encarna al hombre cuya pequñez moral e intelectual no le impide alcanzar el éxito"
- Esta es la razón por la que Val Kilmer no aparece en 'Willow'
- Anne Igartiburu desvela las razones por las que se niega a presentar las Campanadas con Bertín Osborne
- Prohibido especular: España quiere ganar a Japón para cerrar su billete para los octavos
- La opinión de Fernando Hierro sobre los vídeos de TikTok de Iker Casillas: "Eres un ridículo"
- Irán mata a un joven por celebrar la derrota de Irán ante EE UU en el Mundial
- 11 películas de fútbol imprescindibles: la mejor selección de cine para el Mundial de Qatar
- #EmergenciaEnCasa: que la crisis no afecte a la alimentación infantil
- Las acusaciones de Montero al PP soliviantan también al PSOE: "Le ha devuelto el marco a la derecha"
- El hallazgo arqueológico del año en Egipto: retratos de momias, una estatua "rara" y un edificio funerario romano
- **Crimen de Lugo:** El detenido confiesa que mató de 44 puñaladas a Cristina Cabo por no darle dinero
- **Shakira y Piqué firman ante el juez** el acuerdo sobre la custodia de sus hijos
- **El uso de gas para electricidad se disparó en octubre** mientras industria y hogares redujeron su consumo y el de combustible
- Rusia avisa del "enorme" riesgo de un **choque armado entre potencias nucleares**
- Cuarenta de las razas de perro más populares, a fondo
- El PP se concentra frente al Congreso para **exigir a Irene Montero que "rectifique ya"** la ley del 'solo sí es sí'
- **Problemas con el suministro de amoxicilina** en jarabe para niños: fechas previstas de reposición y alternativas recomendadas
- Los beneficios para Alicia Peña del troteo de su marido, Jorge Pérez, con Alba Carrillo
- Varias borrascas atlánticas amenazan un **punto de la Constitución pasado por agua:** estas son las zonas que tendrán más lluvia
- Un campanero de Utrera, censurado de TikTok por grabarse metiendo la cabeza dentro de la campana
- Ignacio Martínez: "Necesitamos más médicos. Ser Hospital Universitario nos permite formar a los profesionales del futuro"
- Cómo negociar la mejor hipoteca en el mercado ante la subida del euríbor
- 5 destinos baratos para viajar por España en diciembre y disfrutar de la Navidad
- La receta barata del día | Frittata de calabacín por menos de 3 euros
- **Segundo cadáver** encontrado en tres días en Barcelona: lo hallan en un parque
- El emotivo mensaje de un padre a los médicos que salvaron a su hijo: "Me han dado la vida"
- La magia de la Navidad, de un mercadillo en Salzburgo a una granja del Tirol
- VÍDEO | Nos llevamos a Lorena Gómez a un parque de bolas: "Me ofrecen ir a Eurovisión cada año, pero no quiero más concursos y jurados"
- A un año de las elecciones
- El imperio de Schrödinger
- Remitente: Ministerio de Igualdad
- La televisión salvada por la paciencia

- PSOE y Unidas Podemos no logran un acuerdo sobre la autodeterminación de los menores en la ley trans y aplazan la disputa dos semanas
- Casi 10.000 menores han expresado conductas suicidas en la última década y solo el 44% ha recibido tratamiento psicológico
- ¿Criptocrash? Miedo a que una quiebra en cadena detrás de FTX devuelva el bitcoin "en un retroceso como no se ha visto jamás"
- Los seguros de salud incluyen los avances de la telemedicina en sus pólizas
- Lorena Castell, ganadora de 'MasterChef Celebrity 7': "Que el programa es un 'show' lo sabemos todos, porque si no sería aburrido"
- Antena 3 sigue liderando la audiencia mientras Telecinco continúa a la baja
- 'El programa de Ana Rosa' visibiliza la nueva polémica del Ministerio de Igualdad: vetan el cartel de una bodega por considerarla sexista
- La tensa conversación entre María Escoté y Patricia Conde: "No dejaste de trabajar por mí"
- Christian Chávez, de 'Rebelde', desvela que su madre le echó de casa cuando le dijo que era gay: "Me dijo que me iba a morir de sida"
- ¿Cuántos años tienen los mejores 'youtubers' españoles?
- Alice Campello, la esposa de Morata, llega a Qatar junto a sus hijos para apoyarlo
- La tensa conversación entre María Escoté y Patricia Conde
- Anne Igartiburu desvela las razones por las que se niega a presentar las Campanadas con Bertín Osborne
- Pedro Sánchez y el Ministerio de Defensa, nuevos objetivos de cartas explosivas
- Frappart, la primera mujer en arbitrar en un Mundial masculino: "Tendré que controlar las emociones"
- España, en la vanguardia internacional de la lucha contra el dopaje
- La española Pilar Lamadrid líder de la nueva disciplina olímpica de vela iQFoil
- El VAR del partido que enfrentará a Canadá y Marruecos dependerá del arbitraje español
- Canelo Álvarez se disculpa con Messi tras amenazarle: "Me dejé llevar por la pasión"
- Ester Expósito ('Venus'): "Me parece absurdo contratar a un actor por el número de seguidores"
- La imposible acrobacia al volante de Anya Taylor-Joy para 'Furiosa' (y eso que no tiene carnet)
- James Cameron anuncia que se retirará tras poner fin a la saga 'Avatar'
- Mucha zarpa y mucha nieve: el tráiler de 'Oso vicioso' ya está aquí
- Los próximos estrenos de series y películas en Amazon Prime Video en diciembre
- Trucos de maquillaje fáciles y rápidos con los que estar perfectas en fiestas, contados por una profesional
- Las "graves consecuencias" que tiene la obesidad sobre la salud de la mujer
- Haz tu parte (por favor)
- Fisioterapeuta, quiropráctico, osteópata... ¿Qué hace cada uno?
- Quique Dacosta: "La patata sería más cara que la trufa negra si la cantidad de producción fuese la misma"
- El restaurante de un pequeño pueblo alavés que ha conseguido su primera estrella Michelin
- El restaurante donde pueden degustarse los platos de la final de MasterChef Celebrity
- Los platos gourmet de Lidl que puedes incluir en tus comidas estas Navidades
- El pueblo más bonito del lago de Como: tiene una cascada secreta entre sus casas
- Cómo se celebran las Navidades en Países Bajos: su Papá Noel viene de España
- Bustarviejo, un campo de trabajo franquista que puede visitarse en Madrid: "Los presos eran mano de obra esclava"
- Opel ya tiene lista la versión cien por cien eléctrica del Astra
- ¿Qué países europeos son los mejores para conducir? España, en el 'top ten'
- Las matriculaciones de turismos y todoterrenos crecen un 10,3% en noviembre
- Josep María Recasens sustituirá a José Vicente de los Mozos al frente de Renault España
- España es el cuarto país con más mujeres inventoras de Europa, pero hay una clara brecha de género

- Redmi Buds 4 y Redmi Buds 4 Pro: unos auriculares Xiaomi de 60 y 100 euros con potente cancelación de ruido
- Google alerta de un nuevo **spyware** de origen español que espía nuestras búsquedas en el navegador de Internet
- Para ahorrar en vivienda: Hacienda tiene una web en la que están todas las casas y pisos que subasta
- Orion sale este viernes de su órbita para comenzar su vuelta a la Tierra
- Casi una de cada dos pacientes con migraña no se diagnostican: "Algunos pasan un infierno"
- Investigadores españoles identifican un marcador para predecir el avance del cáncer de mama
- El motivo por el que tenemos facilidad para recordar lo malo (y aún más lo traumático)
- Una experta revela el secreto para conseguir que un bebé se duerma con facilidad
- Estas son las multas que pueden ponerte si te encuentras dinero en la calle y te lo quedas
- Cómo solicitar el pago fraccionado de impuestos en Alcalá de Henares y beneficiarse de la bonificación del 3%
- Comprueba si tienes derecho a solicitar los descuentos del bono social eléctrico y térmico
- Los jerséis de Navidad no tienen por qué ser feos (¡o sí!): dónde encontrar los mejores modelos para estas fechas
- Manual de las brochas básicas de maquillaje que tienes en casa (¡y puede que uses mal!)
- Los vídeos de limpieza de coches se han viralizado en TikTok: descubre el 'spray' más económico para copiarlos en casa
- Controla la temperatura de todas las estancias de tu casa de forma fácil y por menos de 13 euros gracias a estos mini-termostatos
- El socialista Javier Lambán cree que a España "le habría ido mejor" sin Pedro Sánchez, pero matiza después que cuenta con su lealtad
- Las personas con discapacidad en edad de trabajar y sus pensiones: "Las no contributivas no permiten llevar una vida digna"
- La estación de tren de Alcalá de Henares se reformará con más espacio y nuevos accesos
- BLOG | ¿Cómo se logró proteger al lobo en España? Su gran defensor lo explica en un libro
- Von Braun, el cómic español que descubre al genio nazi tras la llegada del hombre a la Luna
- El perro de Erralla, el enigmático cánido doméstico más antiguo de Europa
- Cae en Madrid una banda dedicada a vender marihuana y hachís a toda Europa
- Anna R. Costa "Uno de los retos de 'Fácil' es demostrar que la discapacidad puede ser arte"
- Mercadona pone a la venta en su perfumería 'Éclat', su nueva colección limitada de belleza para Navidad
- Los niños y niñas de San Ildefonso cantan los números de la pobreza infantil: "400.000 niños se alimentan mal en España"
- Cómo viajar con nuestras mascotas, la importancia de elegir el transportín adecuado para cada ocasión
- Jorge Blass estrena 'Flipar': "El público perdona el error, pero no perdona el aburrimiento"
- Un día de pelos: Una 'tiktokker' se viraliza tras cortarse el pelo y convertirse sin quererlo en Frodo de 'El Señor de los Anillos'
- El ministro de Exteriores, positivo en covid-19 tras regresar de un viaje a Polonia
- Desarticulada una organización criminal que introducía en el mercado carne equina no apta para el consumo humano
- Requisitos para solicitar la ayuda al alquiler en Navarra de hasta 250 euros mensuales: cómo solicitar y plazos
- 20am: lo que necesitas saber para comenzar el día informado
- Buenas Migas: un boletín para los amantes de la gastronomía
- Cinemanía: las mejores recomendaciones de series y películas

- Mundial 20M: todo sobre la Copa del Mundo gratis en tu correo electrónico</a>
- Última hora</a>
- El calendario</a>
- Vídeo en directo</a>
- Las predicciones</a>
- La 'newsletter'</a>
- La policía interceptó en La Moncloa un sobre con material pirotécnico dirigido a Sánchez</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Superviviente de una mara hondureña: "No me enteré de que a mi hija la violaron a punta de pistola hasta que llegué a España"</a>
- Cómo detectar una crisis de ansiedad y qué hacer</a>
- 'As bestas', de Rodrigo Sorogoyen, lidera la carrera a los Goya con 17 nominaciones</a>
- Empecinamiento sobre Melilla </a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Universo Mundial | Fútbol Sapiens. Siga la programación en directo</a>
- Joshua Kimmich bajo vigilancia</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- ¿Qué opciones tiene cada selección de pasar a octavos?</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Vox torpedea al PP y se presenta en su concentración contra Irene Montero sin estar invitado</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Las 72 horas de crisis en Ciudadanos: Villacís sufre la primera gran baja en Madrid y Almeida se frota las manos</a>
- El desafío de envejecer con VIH: "Las personas de largo recorrido sufrimos depresión y ansiedad"</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Elon Musk asegura que en seis meses se implantará el primer chip en un cerebro humano con Neuralink</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- La SER se consolida como la radio más escuchada con 4.161.000 oyentes</a>
- Carlos Navarro, 'El Yoyas', en busca y captura para ingresar en prisión por maltratar a su expareja e hijos</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- 82 ultras del Atlético y el Depor se sientan en el banquillo por la batalla campal en la que murió Jimmy</a>
- La empresa detrás de Conga, el robot aspirador que revolucionó los hogares españoles y que hoy factura 300 millones de euros</a>
- El tiempo en el puente de diciembre: primero frío y, después, lluvias generalizadas</a>
- Shakira y Piqué sellan su acuerdo de separación en Barcelona ante el juez</a>
- Estas son las mejores series de diciembre</a>
- Aló Comidista: "Si dejas la piña boca abajo, ¿se pone más dulce?"</a>
- Por qué un yogur aún hace dudar a muchos españoles</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- México se queda a un gol del pase </a>
- Argentina se gana los octavos </a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Los vídeos de los goles y el resumen de la jornada 11</a>
- Catarí que te vi: cómo llenar el depósito de gasolina por 22 euros en Qatar</a>
- Ramón Besa, periodista de EL PAÍS: "Nadie juega como España"</a>
- Busquets: "Me sorprendió el talento de Gavi y Pedri cuando llegaron al equipo"</a>
- La historia de una niña que soñó con ser pianista</a>

- Pon nombre a lo que te pasa para sentirte mejor</a>
- Volver a empezar</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- ‘Mujeres de armas tomar’: la historia de la víctima, que desesperada por justicia, se vengó de sus violadores</a>
- 100 euros por un caballo viejo para hacerlo filetes: desarticulada una red que producía carne no apta para consumo</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Estos picos han enamorado al chef José Andrés y están hasta en las mesas de Corea del Sur</a>
- La primera simulación cuántica de un agujero de gusano abre una nueva puerta para entender el universo</a>
- Del referéndum de Cataluña a #Cuéntalo: los archiveros guardan tuits que han hecho historia</a>
- Ninguneo administrativo al ciudadano: recibos del IBI que no llegan, llamadas sin respuesta y fallos técnicos</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- La ONU advierte de que una de cada 23 personas requerirá ayuda humanitaria en 2023</a>
- ‘Podcast’ | China pide libertad, no democracia</a>
- Videonálisis | Así pueden afectar las protestas al Gobierno chino</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- ¿Qué beneficios aportan los alimentos húmedos a las mascotas?</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Muere Christine McVie, cantante y teclista de la legendaria banda Fleetwood Mac</a>
- Club Welnia, el programa que nos acompaña en todo momento</a>
- Al norte del Norte: por la remota Ruta de la Costa Ártica de Islandia </a>
- Vídeo | La canción de Antonio Carmona a su padre sobre el alzhéimer: “La escuchó antes de morir”</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Glashütte, el pueblo alemán convertido en capital mundial de los relojes con denominación de origen</a>
- El papel de Marruecos y España en las muertes del 24-J</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- La respuesta de la primera ministra de Nueva Zelanda a una pregunta machista: “¿Os reunís por tener una edad parecida?”</a>
- Irene Montero, al PP en el Congreso: “Ustedes promueven la cultura de la violación”</a>
- ‘Tu opinión es una mierda’: la campaña que alerta sobre los peligros de la polarización</a>
- El propósito empresarial, una seña de identidad</a>
- Las ciudades españolas ganan en salud respecto al año pasado</a>
- Alimentación, empresa y deporte se unen en el pódium de la competitividad</a>
- Nicolás Maduro condiciona la celebración de elecciones libres al levantamiento de sanciones </a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Yolanda Díaz apuesta por “tender la mano” al Gobierno mexicano</a>
- Prisa presenta el plan para redoblar su apuesta por la sostenibilidad</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Arabia Saudí encarga otros cinco buques de guerra a Navantia </a>
- PSOE y Unidas Podemos aplazan el debate sobre la autodeterminación de género en la ‘ley trans’</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Cataluña afronta el invierno de la crisis energética con solo 42 contadores sociales

s de luz</a>

- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- ‘Phishing’, ‘malware’ o ‘ransomware’: el reto de formar en ciberseguridad tras la p

andemia</a>

- La prohibición del maíz transgénico abre un nuevo foco de tensión entre EE UU y Méx

ico</a>

- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- 75 años de la muerte de G.H. Hardy, un matemático excéntrico y brillante</a>
- Hallan en Transilvania una nueva especie de dinosaurio enano herbívoro que vivió ha

ce 70 millones de años</a>

- Un fármaco experimental contra el alzhéimer confirma efectos positivos, pero puede

estar detrás de la muerte de dos pacientes </a>

- Ronna, ronto, quetta y quecto, los nuevos prefijos para magnitudes extraordinarias

</a>

- La Policía de San Francisco contará con robots con capacidad de matar</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Pelé se mantiene estable tras ser hospitalizado de manera inesperada en São Paulo</

a>

- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- El Sónar 2023 anuncia a Richie Hawtin, Laurent Garnier o The Blessed Madonna entre

sus primeros artistas invitados</a>

- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Borja Cobeaga: “Tardé más en sacarme el carné que en hacer esta serie”</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Radiografía de ‘El hormiguero’: entretenimiento blanco de éxito... y polémicas por ma

chismo</a>

- Buckingham fuerza la dimisión de una asistente de la reina Camila por sus comentari

os racistas</a>

- Los príncipes de Gales emulan en su viaje a Boston la visita que la reina Isabel II

hizo a la ciudad en 1976</a>

- Spotify Wrapped, el selfi musical que monopoliza la conversación en Twitter y acari

cia el ego</a>

- La economía solidaria: un nuevo modelo financiero contra la desigualdad y la emerge

ncia climática</a>

- Alima tiene bronquiolitis, pero no hay oxígeno para evitar su asfixia en el hospita

l de Etiopía</a>

- Muchos hombres, una sola mujer</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- El lamentable caso de Anne Hathaway o por qué se odia a algunas famosas sin justifi

cación ninguna</a>

- Samantha Hudson: “Me va bien económicamente, pero no soy Paula Echevarría”</a>
- “Tobogán de piojos”, “genocida de oxígeno”... Por qué cada Mundial recuerda el pode

río del español como idioma para insultar</a>

- Alejandro Jato, el actor que será Camilo Sesto: “Da pena la imagen que los jóvenes

tienen de él, con lo grande que fue”</a>

- Una apasionante ruta por el patrimonio español en peligro de desaparecer</a>
- Nuevos estadios, edificios de premios Pritzker y los zocos de Doha: Qatar, sede de

la arquitectura mundial</a>

- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Descansar está mal visto: el arte perdido del reposo</a>
- Madrid, la nueva Miami: así se han hecho con la capital los ricos latinoamericanos

</a>

- Si no tiene ahorros, esta es una de las pocas alternativas para comprar casa</a>
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- <span class="c\_t\_i c\_t\_i-s \_pr" name="elpais\_ico
- Cómo será el descuento de 20 céntimos en la gasolina en 2023</a>
- El coche que se anticipa al declive de los SUV </a>
- Lentejas con gambas</a>

- ¿Son saludables todas las conservas de verduras y legumbres?
- Francesc Cambó i Jordi Pujol davant del mirall
- En teoria, és progressista
- Pilar Bonet dedica a los periodistas que cubren la guerra de Ucrania el premio Francisco Cerecedo de periodismo
- Taller de 'podcast' narrativo
- La cubertería dorada ideal para esta Navidad: elegante, de 30 piezas y en oferta
- Los mejores termostatos inteligentes último modelo con los que ahorrar en calefacción
- 11 protectores autoadhesivos de paredes y esquinas del garaje para aparcar tranquilo
- Siete gorros de invierno para hombre por menos de 20 euros
- Este árbol de Navidad, que arrasa en Amazon, tiene efecto brillo y se vende en cuatro tamaños
- Hildalgo pide más dinero por Air Europa y retrasa la venta
- Los satélites de Vox: cómo la ultraderecha utiliza una red de asociaciones para aparentar apoyo social a sus postulados
- El demolidor minuto y medio de Batxi López contra Ayuso que aún resuena en los pasillos

In [4]:

```
#Generación de Lista de signos de puntuación

punctuation=[]
for s in string.punctuation:
    punctuation.append(str(s))
sp_punctuation = ["¿", "¡", "“", "”", "...", ":", "-", "»", "«"]

punctuation += sp_punctuation

punctuation[:10] #Ejemplo de Los símbolos de puntuación que están incluidos
```

Out[4]: ['!', '"', '#', '\$', '%', '&amp;', "'", '(', ')', '\*']

In [140]:

```
#Listado de palabras que queremos eliminar del texto
#Es un proceso iterativo por lo que si después vemos que nos siguen quedando "caracteres"
#Existe librerías y listados de "Stop_words", pero por ahora vamos a dejarlo vacío

nltk.download('stopwords') #La primera vez debemos descargar las "stopwords"

stop_words = stopwords.words('spanish') #Listado de palabras a eliminar

stop_words += ["\u200b", "\xa0", "españa", "español", "personas", "bajo", "primera",

stop_words[:10]
```

```
[nltk_data] Downloading package stopwords to
[nltk_data] C:\Users\Alumne_mati1\AppData\Roaming\nltk_data...
[nltk_data] Package stopwords is already up-to-date!
```

Out[140]: ['de', 'la', 'que', 'el', 'en', 'y', 'a', 'los', 'del', 'se']



In [141...

```
#Reemplazamos signos de puntuación por "":

for p in punctuation:
    clean_texto = clean_texto.replace(p, " ")

#Eliminamos espacios blancos, saltos de línea, tabuladores, etc
clean_texto = " ".join(clean_texto.split())

#Reemplazamos stop_words por "":
for stop in stop_words:
    clean_texto_list = clean_texto.split()
    clean_texto_list = [i.strip() for i in clean_texto_list]
    try:
        while stop in clean_texto_list: clean_texto_list.remove(stop)
    except:
        print("Error")
        pass
    clean_texto= " ".join(clean_texto_list)

print('Cantidad de palabras en la cadena "clean_texto":', len(clean_texto))
```

Cantidad de palabras en la cadena "clean\_texto": 13286

In [142...

```
lista_texto = clean_texto.split(" ")

palabras = []

#Paso intermedio para eliminar palabras muy cortas (emoticonos,etc) y muy largas (ulr

for palabra in lista_texto:
    if (len(palabra)>=3 and len(palabra)<18):
        palabras.append(palabra)

print('Cantidad de palabras en la lista "palabras":', len(palabras))
```

Cantidad de palabras en la lista "palabras": 1541

In [143...

```
#Generamos un diccionario para contabilizar las palabras:

word_count={}

for palabra in palabras:
    if palabra in word_count.keys():
        word_count[palabra][0]+=1
    else:
        word_count[palabra]=[1]
```

In [144...

```
#Generamos el DF y Lo ordenamos:

df = pd.DataFrame.from_dict(word_count).transpose()
df.columns=["freq"]
df.sort_values(["freq"], ascending=False, inplace=True)
df.head(15)
```

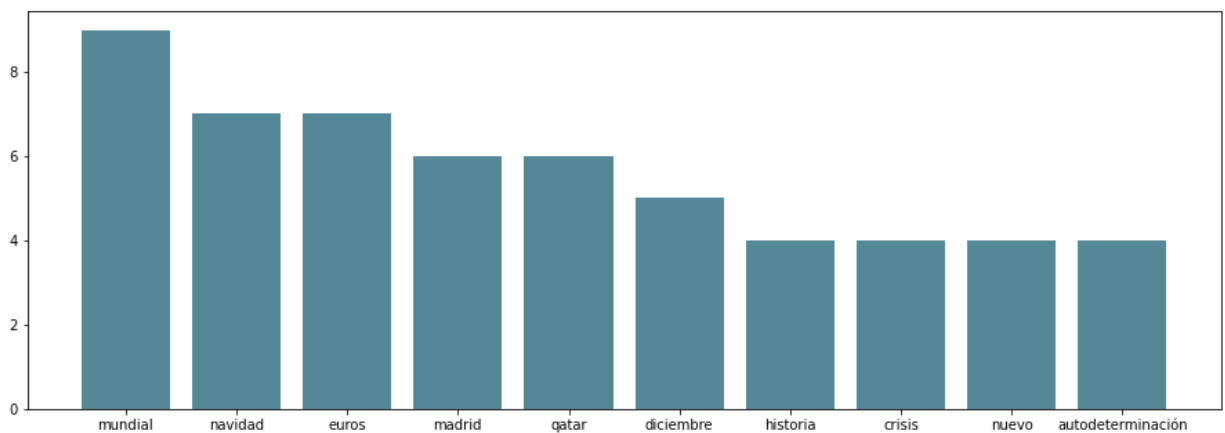
Out [144...

	freq
<b>mundial</b>	9
<b>navidad</b>	7
<b>euros</b>	7
<b>madrid</b>	6
<b>qatar</b>	6
<b>diciembre</b>	5
<b>historia</b>	4
<b>crisis</b>	4
<b>nuevo</b>	4
<b>autodeterminación</b>	4
<b>género</b>	4
<b>millones</b>	4
<b>casa</b>	4
<b>precio</b>	3
<b>modelo</b>	3

Representa el DataFrame de frecuencias como un grafico de barras

In [145...

```
#Muestra de los resultados con un grafico de barras:  
def plot_bar(data=df, top=5):  
    fig = plt.figure()  
    ax = fig.add_axes([0,0,2,1])  
    ax.bar(x =df.iloc[:top,:].index, height = df.iloc[:top,0].values, color="#538897"  
    plt.show()  
  
#Graficamos el TOP 10 palabras por frecuencia  
plot_bar(data=df, top=10)
```



## Muestra las palabras con el módulo WordCloud

In [146...

```
# Transformamos La Lista final en una cadena, para utilizarla con el WordCloud
texto_limpio = " ".join(palabras)

word_cloud = WordCloud(height=800, width=800, background_color='white', max_words=150)

word_cloud.to_file("./img/ejemplo_sencillo.png") #Guardamos La imagen generada

plt.figure(figsize=(10,8))
plt.imshow(word_cloud)
plt.axis('off')
plt.tight_layout(pad=0)
plt.show()
```



### WordCloud con forma:

In [147...

```
#Función para transformar todas las imágenes PNG con fondo transparente a fondo blanco

def transform_white_background(png_path):
    picture = Image.open(png_path).convert("RGBA")
    image = Image.new("RGB", picture.size, "WHITE")
    image.paste(picture, (0, 0), picture)

    plt.imshow(image)

    mask = np.array(image)

    return mask
```

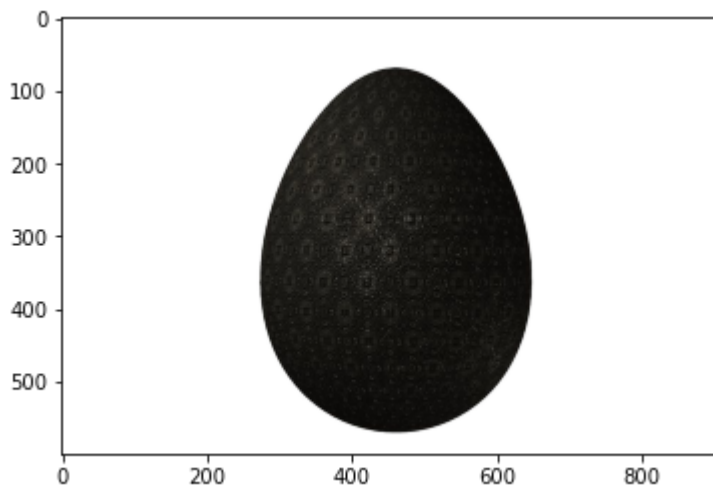
In [148...

```
#En caso de que NO tengas que realizar este paso intermedio, directamente creamos la

image = Image.open("./img/egg.png") #Aquí el path

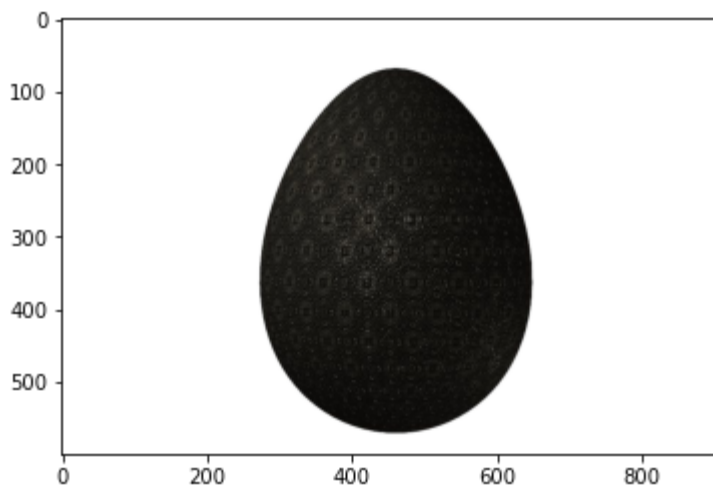
plt.imshow(image)

mask = np.array(image)
```



In [149...

```
mask = transform_white_background("./img/egg.png")
```



```
word_cloud = WordCloud(mask=mask, background_color='white', contour_width=1, contour_
word_cloud.to_file("./img/egg_word_cloud.png") #Guardamos la imagen generada

plt.figure(figsize=(20,10))
plt.imshow(word_cloud)
plt.axis('off')
plt.tight_layout(pad=0)
plt.show()
```

