

nlp_tp_ln

December 5, 2019

1 Ejercicio Procesamiento del Lenguaje Natural

1.1 Analizaremos los titulares del diario La Nacion

2 Que vamos a hacer:

1 - Cargar los datos

- 2 - Limpiar datos
- 3 - Analisis Exploratorio
- 4 - Anáisis de Sentimiento

```
[1]: %%time
# importamos librerias utiles
import requests
import pickle
from time import sleep
```

CPU times: user 207 ms, sys: 59.5 ms, total: 267 ms

Wall time: 1.13 s

2.1 La Nacion

3 1 - Cargar los Datos

```
[2]: %%time
# lo metemos en un Panda's dataframe
import pandas as pd
pd.set_option('max_colwidth',150)
import pickle
# Cargamos el csv de los titulares de La Nacion en un dataframe

data_ln = pd.read_csv('ln_por_year.csv',index_col=0 )
data_ln.index.name=None
data_ln
```

CPU times: user 885 ms, sys: 214 ms, total: 1.1 s
Wall time: 4.15 s

[2]:

Texto

2006 bajo el signo del terror apple renovó toda la familia ipod por las
canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector..
2007 cómo saber si tu computadora será capaz de usar windows vista sin
inconvenientes juegos que acaban de llegar y algunos próximos estrenos la
impor..
2008 cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda
la música y los celulares el futuro de nuestra lengua bases del concur..
2009 la mejor publicidad del año episodio ii algo sobre los vascos mundo
extremo un recorrido por los tesoros culturales de la argentina el largo cami..
2010 el inicio del dakar fue a pura emoción en buenos aires un año para
consolidar el crecimiento parte de los festejos por el bicentenario gritos y
su..
2011 dakar mendoza aporta participantes bajo el calor de la admiración cómo
manejarse con la nueva ortografía ayer fue censurado pero hoy es válid..
2012 concurso el album de la argentina bases víctimas erradas en la masacre de
rosario la inflación hace bajar el ingreso de turistas extranjeros la l..
2013 agreden a una chica filman la paliza y suben el video a las redes
sociales actualidad críticas puerta abierta a la escuela del futuro hay
escrit..
2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en
la playa nervo mujeres bellas y de beats fuertes navidad junto al lago..
2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del
de las ventas y pagos online se hará desde un móvil versos peregrinos..
2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de
medalla tiene la argentina en río la última etapa de las pelucas qué podem..

[3]:

```
%%time  
# Veamos uno de los contenidos  
data_ln.Texto.loc[2007]
```

CPU times: user 1.04 ms, sys: 42 µs, total: 1.09 ms
Wall time: 14.5 ms

[3]:

'cómo saber si tu computadora será capaz de usar windows vista sin
inconvenientes juegos que acaban de llegar y algunos próximos estrenos la
importancia de nivelar hacia arriba un buen año para los libros de comunicación
empresarial polémico cambio de metodología del indec breves polideportivo
preparados para la acción catálogo con aroma veraniego best sellers la vuelta de
la vida transparente feedback el festival de jazz tuvo un cierre deslucido
creció un la inversión publicitaria cartas de lectores gaudio empezó el año
con una derrota ante ferrer el reencuentro con dos grandes la península cada
vez más globalizada variedades best sellers un foro de nuevos medios auspiciado
por lanacion com cartas de lectores el rol de los brujos sigue creciendo
internet sos vos pocos cambios en el ranking de la fifa ya hay uno menos en

gran hermano tormenta diplomática por el gran hermano británico se vienen los rusos cartas de lectores bolívar ganó en el duelo de punteros best sellers jorge rojas brilló en la apertura la papisa juana no fue la única un encuentro de figuras del folklore sharapova avanza sin contratiempos apostillas postales coscoínas cd de clásica una reina con el glamour de versalles los sonidos del nuevo folk y la electrónica un bombardero ante nadal cosquín una caja de sorpresas planeta jazz raptan y usan como escudo humano a un empresario preparan un nuevo paseo de artesanos en palermo viejo lo fusilaron por lo poco que tenía para robarle ee uu insiste en el unilateralismo ensayos de un balance best sellers el combate contra el fantasma del fracaso el boletín de calificaciones que entrega el foro económico mundial ubica a la argentina en el puesto entre países esto da lugar a comparaciones que duelen y asombran en la trastienda andrés romero logró un valioso séptimo lugar videocámaras de última generación en el mercado la banda capital de la chacarera nueve lunas en la balanza intensifican la limpieza en las zonas más sucias de la ciudad la polución vial daña los pulmones y el corazón gibson nos muestra la pasión del maya levantan la basura sacada fuera de hora cartas de lectores best sellers polideportivo cine argentino las opciones de windows vista para el hogar y la familia por doce días la danza moverá a la ciudad mooner vuelve a tener la montura en el lomo detroit es el líder y delfino sigue en alza principio de acuerdo entre hamas y al fatah kirchner desligó el retroceso de los bonos de la polémica por el indec una noche de distintas poéticas señal de largada para cosquín rock encuentros de borges con dios manipular desde la pura belleza el supermercado de la maquinaria el rock echó a andar otro cosquín todas las peñas todas el equipo de seven sigue sin mejorar best sellers pastillas clásicos y hits para todos los gustos brasil ganador en punta en villa maría una fiesta muy popular fiestas y contrastes del rock argentino la triste letra de un sinsentido qué pasa comienza el noveno buenos aires tango la compactera best sellers desilusiones y pocas sorpresas el tiempo de los luchadores queremos explorar otro territorio breves cosechadoras de alta potencia y capacidad de trabajo correo de lectores best sellers el msn messenger sin secretos preguntas frecuentes y consejos útiles tiger woods da clases a domicilio variedades breves a un clic el golf se juega en un gran hermano best sellers amenaza para royal arsenal se quedó con las manos vacías radiografías el zoom precios sostenidos para los granos masiva marcha y choques en montevideo el golpe de acasuso el efecto soja sobre la tv best sellers el aura de la infancia catamarca dio a kirchner el primer triunfo del año nalbandian y chela cartas de la legión para los octavos el silencio no es salud afirma una campaña contra la droga agenda la rioja maza resiste la suspensión nadie ni nada detiene la gran serie triunfal de los spurs la cebit abre sus puertas una película argentina premiada en Málaga el cine del mundo en días explosión de tecnología best sellers intenso itinerario una película española ganó el astor de oro nadal consiguió un título muy esperado las argentinas en el puesto ° del ranking emprendedor cine al día el último libro de harry potter será el segundo en extensión superficial relato sobre vikingos en américa el nuevo flamenco invade a buenos aires guerra entre el incaa y el bafici alejandro sanz en river del potro se dio un gusto grande el hijo responsable del joyero seis de cada diez empresarios creen que la economía está

mejor que un año atrás y que el desempleo seguirá bajando pero tienen un optimismo matizado esta película creo que ya la vi una feria de oportunidades el programa que dio vida a king kong temporal apagón y protesta en la capital chela extiende su buen momento el desafío de crear nuevos lectores la cebit con mayor presencia argentina usurpan tierras vecinas a la villa valora la real academia los mensajes de texto estilo pilar abre sus puertas el cielo la piedra y la justa proporción entre el matadero y los dioses zoom festival internacional de cine de mar del plata un placebo para almas oscuras phelps el gran protagonista el esfuerzo de cañas no alcanzó con djokovic hacia otra visión de la lengua española la casa rosada infló la causa malvinense con aire de campaña desventuras de un país seseante doce días de cine independiente las campañas electorales ganan en agresividad y transgresión a un clic rotter de berlín al abasto proyecciones debates y presentaciones best sellers venezuela hoy la cultura en tiempos de Chávez san pablo la hora del compromiso los que queman los libros bases del concurso rincón gaucho en la escuela fidel castro redobla su actividad en los medios lanacion.com recibió una mención de honor festival de documentales ¿qué le preguntaría a jorge drexler con más de personas comenzó ayer el quilmes rock el superclásico por dentro balance positivo para la fiesta del cine best sellers cancún zona de obras el quilmes rock brilló con los piojos los estudiantes un blanco para quienes odian a la sociedad distinguen a la nación en internet polideportivo creció un la industria por los autos y la construcción libros ensayos de lunes a viernes cacerolazo en río gallegos para apoyar el reclamo docente minuciosa biografía sin alma final cut studio a todo color premio a la excelencia exportadora la excelencia se renueva crece arteba y apuesta al mercado la vuelta al cine de sergio renán sólo para milongueros convalidaron la excarcelación de una mula de drogas no fue el destino polideportivo libros el under del siglo veintiuno la tercera medida de fuerza en diez meses tv y radio están nominados camino a un nodo logístico una serie de documentos decisivos bajo la lupa ee uu vigilará la piratería el dorrego vuelve el salón del automóvil en un año con producción récord retrato imaginario de una fotógrafa famosa la fobia de los nominados un directivo de techint en la casa rosada los latinos más bellos según la revista people yukos devorada por sus competidores a un clic nik una atracción singular en la muestra al borde de un acuerdo por la carne carbap ratificó su paro la alianza del arte y la solidaridad pidieron el procesamiento de directivos de southern winds no te va gustar llega a obras ocio negocios recelos por la venta de drogas ilícitas cartas de lectores empezaría a recuperarse la oferta ganadera un domingo a pedido de gran hermano famosos el gran hermano farandulero temen en chile que los cortes de gas llegue a los hogares roma campeón de la copa italia jarvis mezcla a motorhead con disney se prepara san sebastián en la trastienda hoy última oportunidad para asistir a expo nieve reunión cumbre para homenajear al cine a la espera de harry potter bases del concurso rincón gaucho en la escuela nuevos negocios con australia diez desarrolladores argentinos buscarán inversores en España expectativas y controversias la radio también se hizo escuchar una vidriera para hollywood protagonistas de la experiencia tengo un partido difícil con del potro una vidriera se abre al mundo la guía definitiva para conocer todos los formatos de

memorias flash actuales un panorama muy positivo el libro y el viaje que hicieron historia armas garré admitió irregularidades estilo diva el cuarteto de nos arrasó en los graffiti placeres sobre ruedas los tenistas íntimos en la red kuitca por dos en venecia el voto obligatorio puede ser una trampa para esclavizarnos se pagaron por un toro braford otra joya de un talentoso director variedades favorecer los talentos diálogo imaginario con un lector de hoy monitor polideportivo el nuevo león de venecia cartas de lectores poner el cuerpo y la imaginación shrek vuelve con brillo pero con una trama menos eficaz paso en falso de una joven promesa espacio ferroviario para la muestra fiabci y el ciclo de conferencias una trilogía de Raúl Perrone los ¿década maldita vida empresaria el salón sintetiza la decisión de apoyar la recuperación industrial la fiesta interminable del pop en pantalla en las estaciones de servicio no hubo una acción uniforme con los precios los créditos más cerca historias en carne propia alemania críticas por la caricatura de merkel proponen medidas de ahorro en la ciudad premian al escritor israelí amos oz mundo xtremo prensa y poder el imperio murdoch bergoglio equiparó el aborto con el parricidio breves premios al periodismo solidario una automotriz renovó su eslogan con una amplia campaña publicitaria un teatro que recupera la música cartas de lectores un matadero mediático brasil sigue sin convencer juego ecuestre para los ejecutivos el dólar seguirá estable según los resultados del barómetro de empresas que realiza la consultora y auditora deloitte la visión es que hay cautela entre los empresarios el ingreso en la villa fue el lado débil de la seguridad al rescate de la milonga candombe cuál es el límite del mercado estamos todos narcotizados mirando a través del ojo de la cerradura la vida de otros paso en falso del joven mago harry potter se adelantó en giffoni harry potter el último hechizo del mago un mundo de libros abrió sus puertas y espera a los chicos dulce cosecha en el valle inferior del río chubut cartas de lectores distinguen a las pymes exportadoras el artista que desafía a la industria versión elemental de un personaje famoso la crisis energética golpeó fuerte a ocho industrias una fiesta global recibió a harry potter el futuro desde el corazón de hp la vida el rock'n roll y todo lo demás la carrera contra los ladrones de arte james baldwin y estambul todo el año puede ser virtual con nombre y apellido una pluralidad de plurales primeras experiencias con el intel core duo desventuras de campesinos los monólogos tienen su fiesta otro récord de harry potter vendió libros por minuto españa rectifican la polémica caricatura con la mirada en el siglo xxi el dorrego la cálida nostalgia de un amor que no fue la fiesta que no oculta la crisis la premiación nuestra de cada día un viaje que en la intimidad mostró a un gobierno bicéfalo falta poco para que comience la xi bienal porteña insiste y serás un gran hermano psicodélico vuelve al ruedo más sólida la gran semana una nevada tiñó de blanco a tucumán cartas de lectores la identidad como motor del cambio un proyecto con la mirada puesta en el futuro casares un testigo de la despedida de borges los simpson dueños de las salas documentales políticos de estreno el caño pudo más telecom cuadruplicó sus ganancias encuentro en nueva york aftosa continúa la lucha en busca de la propia extinción un clásico con todo su contexto gritos y susurros en los ee uu creen que el caso afectará a cristina kirchner se acercan las estrellas paisajes y sonidos bellos el tiempo le sienta bien a willis jorge

rojas trae sus novedades la discapacidad en la era digital películas de montaña topografía de una literatura la venganza de kerouac el gran simulador bayern golea en el clásico a un clic buen debut de acasuso en new haven rescatan a un niño de las aguas heladas usina para el arte nuevo empresas alto vuelo repsol ypf esso y petrobras afirman que retrotraerán los aumentos el libro y el viaje que hicieron historia una lágrima para el negro la crisis energética golpeó fuerte a la industria en julio el príncipe que anticipó el sionismo música y cine según pasan los años música y cine según pasan los años poblar y purificar el otro ramón carrillo la argentina consolida su juego y se prepara para la etapa decisiva una muestra con historia para respetar la ortografía a veces hay que cambiarla final a toda orquesta en el quinto mundial de tango breves rally de las montañas las razones del éxito de un equipo que se sobrepuso a las ausencias la carrera contra los ladrones de arte adelantando el futuro prêtàthonet confidencias extremas van por el buen camino golf en todas partes la vidriera agenda remate de arte solidario llega a millones la inversión en promoción en puntos de compra bafweek con los chemical a la cabeza crece veces más la esperanza de vida en occidente greenspan ve características idénticas con crisis anteriores el lado desconocido de algunas canciones tender la mano un pequeño nuevo ancestro de las aves abra pampa prepara la feria de la llama ¿las apariencias engañan política y verdad larga la maratón de festivales a un clic nueve periodistas de la nacion reciben el konex de platino el salón náutico a pleno nueve konex de platino para la nacion el best seller que volvió de la muerte cine argentino yo me quiero casar i un pasado que pudo ser posible dioses aztecas en el barrio de núñez los effie awards distinguieron las campañas más efectivas los extranjeros fueron las estrellas cartas de lectores cartas de lectores el último tango en finlandia bajó el telón en ushuaia último viaje del peregrino inmóvil casa foa recibió a los primeros huéspedes de un viaje por éxitos de la canción italiana la semana del arte a toda música de cómo la borraja perdió su buena fama brillante inauguración de una semana musical fantasías y verdades sobre los incunables apenas una batería de efectos especiales hasta la próxima primavera un curso con representantes de países hoy el gran cierre de la semana del arte la compactera el ganador del nobel de literatura ya fue designado los secretos detrás de la elección del premio nobel de literatura cien volando síntesis en las tribunas se jugó el partido de las provocaciones se requiere inexperiencia también celebro la hermosura del mundo la vanguardia en londres manu chao está loco por la música europa atrae a las empresas argentinas una idea en el pedestal avanza la feria de galerías abrió sus puertas la feria del libro de francfort según los resultados del barómetro de empresas que elabora la auditora y consultora deloitte los hombres de negocios creen que se deterioran las expectativas inflación y salarios entre las preocupaciones para seguir entreteniéndose para dar una mano diálogo con alfredo casero el más exiliado de los exiliados la revancha de las industrias culturales impulsada por las automotrices la industria creció dos bienales en dos semanas memorables danza en el mundo música popular servicios de apple para programadores ganó carrió en los estados unidos el fin de la rotación agüero se supera cada día cantando por un sueño culminó el circuito audi la semana que pasó cine argentino cartas de lectores los gaturros suman otra

atracción la raspadita virtual soda stereo piden una prórroga de la moratoria de la ciudad reclamo de mayor inversión energética chicago recuperó la motivación y triunfó gran protagonismo de los semilleros en expoagro distinción para alterini se corrieron los nike aires electrónicos y refrigerantes fabián casas distinguido ginóbili sigue imparabile sumó un punto por minuto hacelo corto catálogo de libros tres unipersonales muy diferentes fiesta de los clásicos más antiguos del país retenciones el ruralismo duro apuesta al diálogo una invitación a dominar el tiempo interesante exponente del cine de acción volvieron a aumentar los combustibles novedades de las empresas en primer plano ediciones de tecnología amigos a contrapelo las primeras ediciones del suplemento más de once años reflejando la vida digital gran interés por la feria internacional de turismo harry potter vuelve a hablar castellano harry potter con fecha y hora exactas breves grandes relatos aborígenes esta vez el gusto se lo dio sampras para el programa de este año recibieron currículum uruguay cortó otro puente y se agudiza el conflicto libros y revistas billete de magneto con el de descuento delfino ganó el duelo ante nocioni cartas de lectores más de artistas se lucirán en san isidro nueva aproximación a un crimen el nuevo enemigo de batman atractiva muestra de tecnología en palermo gelman obtuvo el premio cervantes concursos de novela y ensayo bases del concurso la nacion sudamericana de novela y ensayo ¿qué es ser culto hoy secuencias narrativas recuerdos de la crisis de américa latina bajo la lupa europea tejiendo nuevas redes cine argentino para dar una mano rottemberg mediador en el conflicto de la tv un mundo de creación disfrutable experiencia preservarán el bajo belgrano cartas de lectores clásica para todos en la puna esquilan vicuñas silvestres la nacion ganadera se hará en tandil breves la dura advertencia a vázquez copó la prensa uruguaya los tenistas no se callan nada largo camino al oscar maceo parker el saxo del funk cartas de lectores con la copa en alto montaje filosófico se prepara estilo pilar vuelven a prohibir la oferta sexual de travestis en palermo maratón de tango para pasar el invierno terminó el censo de los chicos de la calle publicaciones la grilla de cosquín interés sostenido de la demanda sorpresas hallan un cadáver dentro de una valija en villa crespo la buena calidad sostuvo los precios bases del concurso la nacion sudamericana de novela y ensayo libros argentinos para el nazismo amantes y suicidio a evita la engañaron con el cáncer polideportivo nuevos aires para el festival internacional lipovetsky y la felicidad libros y revistas positivo días sergio wolf el nuevo director del bafici sergio wolf al frente del bafici cartas de lectores canon de libros criollos para argentinos de hoy cuando la imagen es noticia más pruebas comprometen al piquetero que copó un ministerio'

4 3 - Limpiar los Datos

```
[4]: %%time
# Aplicaremos varios rounds de limpieza
import re
import string

def clean_text_round1(text):

    text = text.lower()
    text = re.sub('\[.*?\]\%', ' ', text)
    text = re.sub('[%s]' % re.escape(string.punctuation), ' ', text)
    text = re.sub('\w*\d\w*', '', text)
    return text

round1 = lambda x: clean_text_round1(x)
```

CPU times: user 19 µs, sys: 4 µs, total: 23 µs
Wall time: 32.4 µs

```
[5]: %%time
# vemos la primer limpieza
data_ln_clean = pd.DataFrame(data_ln.Texto.apply(round1))
data_ln_clean
```

CPU times: user 43.4 ms, sys: 11 µs, total: 43.4 ms
Wall time: 43.1 ms

```
[5]:
```

	Texto
2006	bajo el signo del terror apple renovó toda la familia ipod por las canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007	cómo saber si tu computadora será capaz de usar windows vista sin inconvenientes juegos que acaban de llegar y algunos próximos estrenos la impor...
2008	cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda la música y los celulares el futuro de nuestra lengua bases del concur...
2009	la mejor publicidad del año episodio ii algo sobre los vascos mundo extremo un recorrido por los tesoros culturales de la argentina el largo cami...
2010	el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...
2011	dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válido...
2012	concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...
2013	agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay


```

escrit...
2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en
la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...
2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del
de las ventas y pagos online se hará desde un móvil versos peregrinos...
2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de
medalla tiene la argentina en río la última etapa de las pelucas qué podem...

```

```

[6]: %%time
      # Segundo round
      def clean_text_round2(text):

          text = re.sub('[\'\"\"...«»]', '', text)
          text = re.sub('\n', ' ', text)
          return text

      round2 = lambda x: clean_text_round2(x)

```

```

CPU times: user 13 µs, sys: 3 µs, total: 16 µs
Wall time: 25.5 µs

```

```

[7]: %%time
      # veamos como queda
      data_ln_clean = pd.DataFrame(data_ln_clean.Texto.apply(round2))
      data_ln_clean

```

```

CPU times: user 2.79 ms, sys: 0 ns, total: 2.79 ms
Wall time: 2.76 ms

```

```

[7]:                                     Texto
2006 bajo el signo del terror apple renovó toda la familia ipod por las
canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007 cómo saber si tu computadora será capaz de usar windows vista sin
inconvenientes juegos que acaban de llegar y algunos próximos estrenos la
impor...
2008 cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda
la música y los celulares el futuro de nuestra lengua bases del concur...
2009 la mejor publicidad del año episodio ii algo sobre los vascos mundo
extremo un recorrido por los tesoros culturales de la argentina el largo cami...
2010 el inicio del dakar fue a pura emoción en buenos aires un año para
consolidar el crecimiento parte de los festejos por el bicentenario gritos y
su...
2011 dakar mendoza aporta participantes bajo el calor de la admiración cómo
manejarse con la nueva ortografía ayer fue censurado pero hoy es válid...
2012 concurso el album de la argentina bases víctimas erradas en la masacre de
rosario la inflación hace bajar el ingreso de turistas extranjeros la l...
2013 agreden a una chica filman la paliza y suben el video a las redes

```

sociales actualidad críticas puerta abierta a la escuela del futuro hay
 escrit...
 2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en
 la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...
 2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del
 de las ventas y pagos online se hará desde un móvil versos peregrinos...
 2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de
 medalla tiene la argentina en río la última etapa de las pelucas qué podem...

```
[9]: %%time
# Como no tenemos un Lemmatizer en español, hacemos manualmente algunas
→ conversiones
# OJO: esto realmente no se hace a mano!!!

def detectadas(palabra):
    eliminar_s =
    → ('libreros', 'textos', 'papelitos', 'monedas', 'páginas', 'anécdotas', 'perros', 'cuadernos', 'blog
    → 'revistas', 'caballos', 'vecinos', 'madres', 'puntos', 'ricos', 'libros')
    if palabra in eliminar_s :
        return palabra[:-1]
    eliminar_es =
    → ('mundiales', 'lectores', 'campeones', 'maníes', 'ustedes', 'autores')
    if palabra in eliminar_es:
        return palabra[:-2]
    return palabra

def clean_text_round3(text):
    ''' '''
    return " ".join([detectadas(word) for word in text.split()])

round3 = lambda x: clean_text_round3(x)
```

CPU times: user 12 µs, sys: 2 µs, total: 14 µs
 Wall time: 24.3 µs

```
[10]: %%time
#vemos como queda
data_ln_clean = pd.DataFrame(data_ln_clean.Texto.apply(round3))
data_ln_clean
```

CPU times: user 33.9 ms, sys: 4.43 ms, total: 38.3 ms
 Wall time: 37.4 ms

```
[10]: Texto
2006 bajo el signo del terror apple renovó toda la familia ipod por las canchas  

argentinos en francia ¿un loco suelto en hollywood cartas de lector int...
```

2007 cómo saber si tu computadora será capaz de usar windows vista sin inconvenientes juegos que acaban de llegar y algunos próximos estrenos la import...

2008 cartas de lector el misterio de rimbaud en la playa al ritmo de la moda la música y los celulares el futuro de nuestra lengua bases del concurso l...

2009 la mejor publicidad del año episodio ii algo sobre los vascos mundo extremo un recorrido por los tesoros culturales de la argentina el largo camin...

2010 el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...

2011 dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válido la ...

2012 concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la li...

2013 agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay escritor...

2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en la playa nervo mujeres bellas y de beats fuertes navidad junto al lago un ...

2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del de las ventas y pagos online se hará desde un móvil versos peregrinos el ...

2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de medalla tiene la argentina en río la última etapa de las pelucas qué podemos ...

```
[11]: %%time
# Esto es un nuevo campo por si quisieramos agregar alguna info adicional a
→ cada año
# Nuestro caso repetimos los años, nos servirá para alguna visualización
full_names =
→ ['2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015', '2016']

data_ln['full_name'] = full_names
data_ln
```

CPU times: user 1.92 ms, sys: 407 µs, total: 2.33 ms

Wall time: 2.09 ms

```
[11]:                                     Texto \
2006 bajo el signo del terror apple renovó toda la familia ipod por las
canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007 cómo saber si tu computadora será capaz de usar windows vista sin
inconvenientes juegos que acaban de llegar y algunos próximos estrenos la
impor...
2008 cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda
la música y los celulares el futuro de nuestra lengua bases del concur...
2009 la mejor publicidad del año episodio ii algo sobre los vascos mundo
extremo un recorrido por los tesoros culturales de la argentina el largo cami...
```

2010 el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...

2011 dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válid...

2012 concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...

2013 agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay escrit...

2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...

2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del de las ventas y pagos online se hará desde un móvil versos peregrinos...

2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de medalla tiene la argentina en río la última etapa de las pelucas qué podem...

	full_name
2006	2006
2007	2007
2008	2008
2009	2009
2010	2010
2011	2011
2012	2012
2013	2013
2014	2014
2015	2015
2016	2016

```
[12]: %%time
# Hacemos el pickle para usar más adelante
data_ln.to_pickle("corpus_ln.pkl")
```

CPU times: user 2.35 ms, sys: 492 µs, total: 2.84 ms
Wall time: 2.42 ms

```
[13]: %%time
data_ln.Texto[0:255]
```

CPU times: user 858 µs, sys: 182 µs, total: 1.04 ms
Wall time: 1.05 ms

```
[13]: 2006 bajo el signo del terror apple renovó toda la familia ipod por las
canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007 cómo saber si tu computadora será capaz de usar windows vista sin
inconvenientes juegos que acaban de llegar y algunos próximos estrenos la
```

```

impor...
2008    cartas de lectores el misterio de rimbaud en la playa  al ritmo de la
moda  la música y los celulares el futuro de nuestra lengua bases del concur...
2009    la mejor publicidad del año  episodio ii algo sobre los vascos mundo
extremo un recorrido por los tesoros culturales de la argentina el largo cami...
2010    el inicio del dakar fue a pura emoción en buenos aires un año para
consolidar el crecimiento parte de los festejos por el bicentenario gritos y
su...
2011    dakar    mendoza aporta  participantes bajo el calor de la admiración
cómo manejarse con la nueva ortografía ayer fue censurado  pero hoy es válid...
2012    concurso el album de la argentina  bases víctimas erradas en la masacre
de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...
2013    agreden a una chica  filman la paliza y suben el video a las redes
sociales actualidad críticas puerta abierta a la escuela del futuro  hay
escrit...
2014    el niño grande que nos sigue sorprendiendo en punta  la fiesta siguió en
la playa nervo mujeres bellas y de beats  fuertes navidad junto al lago...
2015    sospechan que a lola chomnalez la habrían asesinado el sábado en  más
del  de las ventas y pagos online se hará desde un móvil versos peregrinos...
2016    pinamar  una nueva edición de la maratón del desierto ¿qué chances de
medalla tiene la argentina en río  la última etapa de las pelucas qué podem...
Name: Texto, dtype: object

```

```

[14]: %%time
# Creamos una matriz de palabras y quitamos las palabras mas comunes del
↳ castellano
from sklearn.feature_extraction.text import CountVectorizer

# En spanish.txt tenemos una lista de palabras comunes
with open('spanish.txt') as f:
    lines = f.read().splitlines()

cv = CountVectorizer(stop_words=lines)
data_cv = cv.fit_transform(data_ln_clean.Texto)
data_dtm = pd.DataFrame(data_cv.toarray(), columns=cv.get_feature_names())
data_dtm.index = data_ln_clean.index
data_dtm

```

CPU times: user 582 ms, sys: 172 ms, total: 754 ms
Wall time: 2.81 s

```

[14]:

```

	abandona	abandonar	abandonó	abanico	abasto	abdicación	abdominal	\
2006	0	0	0	0	0	0	0	
2007	0	0	0	0	1	0	0	
2008	0	0	0	1	0	0	0	
2009	0	0	0	0	0	0	0	
2010	0	0	0	0	1	0	0	

2011	0	0	0	0	0	0	0
2012	1	0	1	0	1	0	0
2013	0	0	0	0	0	0	0
2014	0	0	0	0	0	1	1
2015	0	1	0	0	0	0	0
2016	0	0	0	0	0	0	0

	abierta	abiertas	abierto	...	órganos	última	últimamente	último	\
2006	0	0	1	...	0	0	0	0	
2007	0	0	0	...	0	2	0	4	
2008	0	0	0	...	0	0	1	5	
2009	0	1	0	...	0	0	0	2	
2010	0	0	0	...	0	1	0	1	
2011	0	0	1	...	1	1	0	0	
2012	0	0	0	...	0	0	0	0	
2013	2	0	0	...	0	0	0	1	
2014	3	0	1	...	0	0	0	2	
2015	0	0	2	...	0	2	0	1	
2016	0	0	0	...	0	1	0	2	

	últimos	única	único	únicos	útil	útiles
2006	0	0	0	0	0	0
2007	0	1	0	0	0	1
2008	0	1	0	0	0	0
2009	1	0	0	0	0	0
2010	0	0	0	0	0	0
2011	0	0	1	1	1	0
2012	0	0	1	0	0	1
2013	0	0	1	0	0	0
2014	0	0	0	0	0	0
2015	1	0	0	0	0	0
2016	0	0	0	0	0	0

[11 rows x 7334 columns]

```
[15]: %%time
# Lo guardamos como pickle
data_dtm.to_pickle("dtm_ln.pkl")
```

CPU times: user 6.07 ms, sys: 500 µs, total: 6.57 ms
Wall time: 5.9 ms

```
[16]: %%time
# Lo guardamos como pickle también
data_ln_clean.to_pickle('data_ln_clean.pkl')
pickle.dump(cv, open("cv.pkl", "wb"))
```

CPU times: user 34 ms, sys: 7.9 ms, total: 41.9 ms

Wall time: 42.1 ms

5 4 - Análisis Exploratorio

```
[38]: %%time
      # Leemos el documento
      import pandas as pd

      data_ln = pd.read_pickle('dtm_ln.pkl')
      data_ln = data_ln.transpose()
      data_ln.head()
```

CPU times: user 11.8 ms, sys: 3.97 ms, total: 15.8 ms

Wall time: 12.4 ms

```
[38]:
```

	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
abandona	0	0	0	0	0	0	1	0	0	0	0
abandonar	0	0	0	0	0	0	0	0	0	1	0
abandonó	0	0	0	0	0	0	1	0	0	0	0
abanico	0	0	1	0	0	0	0	0	0	0	0
abasto	0	1	0	0	1	0	1	0	0	0	0

```
[39]: %%time
      # Buscamos las palabras mas comunes por año
      top_dict = {}
      for c in data_ln.columns:
          top = data_ln[c].sort_values(ascending=False).head(30)
          top_dict[c]= list(zip(top.index, top.values))

      top_dict
```

CPU times: user 18.3 ms, sys: 160 µs, total: 18.4 ms

Wall time: 17.1 ms

```
[39]: {2006: [('polideportivo', 9),
            ('lector', 7),
            ('cartas', 7),
            ('sellers', 5),
            ('best', 5),
            ('libro', 5),
            ('hoy', 3),
            ('dos', 3),
            ('aires', 3),
            ('buenos', 3),
            ('rural', 3),
            ('mejor', 3),
```

```

('xp', 2),
('controles', 2),
('día', 2),
('crece', 2),
('poder', 2),
('obras', 2),
('premiaron', 2),
('windows', 2),
('salto', 2),
('peso', 2),
('premio', 2),
('recibir', 2),
('país', 2),
('sorín', 2),
('exportador', 2),
('diaz', 2),
('pasa', 2),
('mundo', 2)],
2007: [('lector', 19),
('cartas', 17),
('libro', 15),
('best', 14),
('cine', 14),
('sellers', 13),
('gran', 13),
('nuevo', 8),
('potter', 8),
('harry', 8),
('fiesta', 7),
('mundo', 7),
('arte', 7),
('breves', 7),
('año', 7),
('vida', 7),
('rock', 6),
('música', 6),
('hermano', 6),
('nacion', 6),
('polideportivo', 6),
('hoy', 6),
('argentino', 5),
('viaje', 5),
('crisis', 5),
('semana', 5),
('según', 5),
('vuelve', 5),
('muestra', 5),

```



```

    ('empresas', 5)],
2008: [('lector', 17),
    ('cartas', 15),
    ('concurso', 10),
    ('fiesta', 10),
    ('argentina', 10),
    ('bases', 10),
    ('nuevo', 10),
    ('mejor', 10),
    ('novela', 8),
    ('nueva', 8),
    ('nacion', 8),
    ('cine', 7),
    ('música', 7),
    ('gran', 7),
    ('premio', 7),
    ('polideportivo', 6),
    ('ensayo', 6),
    ('récord', 6),
    ('ganó', 6),
    ('potter', 6),
    ('arte', 6),
    ('año', 6),
    ('noche', 6),
    ('sudamericana', 6),
    ('años', 5),
    ('argentinos', 5),
    ('libro', 5),
    ('feedback', 5),
    ('festival', 5),
    ('publicidad', 5)],
2009: [('agenda', 21),
    ('mejores', 20),
    ('internet', 20),
    ('videos', 17),
    ('insólitos', 17),
    ('argentina', 14),
    ('lector', 12),
    ('mundo', 12),
    ('cartas', 11),
    ('crisis', 11),
    ('nueva', 10),
    ('nuevo', 9),
    ('mejor', 8),
    ('argentinas', 8),
    ('año', 8),
    ('mundial', 8),

```

```

('historia', 7),
('primera', 7),
('gran', 7),
('bases', 7),
('fiesta', 6),
('gobierno', 6),
('gritos', 6),
('conocé', 6),
('premio', 6),
('arte', 6),
('susurros', 6),
('libro', 5),
('vuelve', 5),
('semana', 5)],
2010: [('mejores', 33),
('videos', 26),
('insólitos', 24),
('internet', 23),
('aires', 11),
('buenos', 11),
('cine', 10),
('mundial', 9),
('nuevos', 8),
('gran', 8),
('noche', 8),
('año', 8),
('argentina', 8),
('lector', 8),
('libro', 7),
('argentinos', 7),
('cómo', 7),
('nuevo', 6),
('feedback', 6),
('llega', 6),
('casa', 6),
('mundo', 6),
('polideportivo', 6),
('dakar', 6),
('cartas', 6),
('nueva', 6),
('consumo', 5),
('según', 5),
('días', 5),
('récord', 5)],
2011: [('argentina', 14),
('libro', 11),
('gran', 9),

```

```

('nuevo', 8),
('cinco', 8),
('polideportivo', 8),
('hoy', 8),
('historia', 8),
('común', 7),
('mejor', 7),
('gadgets', 7),
('cómo', 7),
('geekmanía', 7),
('lector', 7),
('vuelve', 6),
('ipad', 6),
('año', 6),
('encuentro', 6),
('feria', 6),
('noche', 6),
('bafweek', 6),
('cartas', 6),
('mundo', 6),
('digital', 5),
('argentino', 5),
('argentinas', 5),
('mundial', 5),
('rs', 5),
('puro', 5),
('tras', 5)],
2012: [('gran', 16),
('nueva', 10),
('argentina', 9),
('fiesta', 9),
('nuevo', 9),
('festival', 9),
('país', 8),
('cómo', 7),
('windows', 7),
('mejor', 7),
('mundo', 7),
('noche', 7),
('obra', 6),
('gobierno', 6),
('casa', 6),
('libro', 6),
('millones', 5),
('lector', 5),
('martín', 5),
('grammy', 5),

```

```

('industria', 5),
('polideportivo', 5),
('equipo', 5),
('microsoft', 5),
('moreno', 5),
('premios', 5),
('semana', 5),
('argentino', 5),
('puerto', 5),
('feedback', 4)],
2013: [('cultural', 17),
('agenda', 17),
('argentina', 15),
('libro', 13),
('feria', 9),
('mundo', 9),
('nueva', 9),
('buenos', 9),
('papa', 8),
('cómo', 8),
('aires', 8),
('nuevo', 7),
('país', 7),
('presentó', 7),
('historia', 6),
('cristina', 6),
('vez', 6),
('festival', 6),
('regreso', 6),
('pantallas', 6),
('ciudad', 6),
('fin', 6),
('mejor', 6),
('dos', 6),
('gobierno', 6),
('diez', 6),
('francisco', 6),
('apple', 6),
('teatro', 5),
('arte', 5)],
2014: [('mundial', 14),
('nuevo', 11),
('argentina', 11),
('fiesta', 10),
('hoy', 10),
('cómo', 9),
('mundo', 9),

```

```

('arte', 9),
('cine', 8),
('aires', 8),
('años', 7),
('brasil', 7),
('libro', 7),
('buenos', 7),
('ganó', 7),
('año', 6),
('boca', 6),
('premio', 6),
('días', 6),
('oscar', 5),
('nueva', 5),
('hot', 5),
('tres', 5),
('cristina', 5),
('tras', 5),
('síntesis', 5),
('rock', 5),
('comienza', 5),
('polideportivo', 5),
('festival', 5)],
2015: [('nueva', 11),
('fin', 10),
('edición', 10),
('argentina', 10),
('mundo', 9),
('año', 9),
('fiesta', 9),
('años', 8),
('gran', 8),
('mejor', 8),
('lector', 8),
('semana', 8),
('síntesis', 7),
('cartas', 7),
('premios', 7),
('the', 7),
('final', 7),
('dos', 6),
('aires', 6),
('buenos', 6),
('argentino', 6),
('feria', 6),
('guía', 5),
('futuro', 5),

```

```

('nuevos', 5),
('dakar', 5),
('vuelve', 5),
('noche', 5),
('tras', 5),
('libro', 5)],
2016: [('lector', 6),
('cartas', 6),
('verano', 3),
('viene', 3),
('nueva', 3),
('fiesta', 3),
('tan', 3),
('muerte', 2),
('dakar', 2),
('negocio', 2),
('foco', 2),
('pudo', 2),
('fin', 2),
('extraño', 2),
('fútbol', 2),
('maratón', 2),
('ces', 2),
('segundo', 2),
('secretos', 2),
('arte', 2),
('ser', 2),
('argentina', 2),
('último', 2),
('terminó', 2),
('punta', 2),
('loeb', 2),
('boca', 2),
('lista', 1),
('libro', 1),
('tv', 1)]]}

```

```

[40]: %%time
# Las mostramos un poco mas ordenadas
for anio, top_words in top_dict.items():
    print(anio)
    print(', '.join([word for word, count in top_words[0:14]]))
    print('---')

```

2006
 polideportivo, lector, cartas, sellers, best, libro, hoy, dos, aires, buenos,
 rural, mejor, xp, controles

```

---
2007
lector, cartas, libro, best, cine, sellers, gran, nuevo, potter, harry, fiesta,
mundo, arte, breves
---
2008
lector, cartas, concurso, fiesta, argentina, bases, nuevo, mejor, novela, nueva,
nacion, cine, música, gran
---
2009
agenda, mejores, internet, videos, insólitos, argentina, lector, mundo, cartas,
crisis, nueva, nuevo, mejor, argentinas
---
2010
mejores, videos, insólitos, internet, aires, buenos, cine, mundial, nuevos,
gran, noche, año, argentina, lector
---
2011
argentina, libro, gran, nuevo, cinco, polideportivo, hoy, historia, común,
mejor, gadgets, cómo, geekmanía, lector
---
2012
gran, nueva, argentina, fiesta, nuevo, festival, país, cómo, windows, mejor,
mundo, noche, obra, gobierno
---
2013
cultural, agenda, argentina, libro, feria, mundo, nueva, buenos, papa, cómo,
aires, nuevo, país, presentó
---
2014
mundial, nuevo, argentina, fiesta, hoy, cómo, mundo, arte, cine, aires, años,
brasil, libro, buenos
---
2015
nueva, fin, edición, argentina, mundo, año, fiesta, años, gran, mejor, lector,
semana, síntesis, cartas
---
2016
lector, cartas, verano, viene, nueva, fiesta, tan, muerte, dakar, negocio, foco,
pudo, fin, extraño
---
CPU times: user 4.32 ms, sys: 24 µs, total: 4.34 ms
Wall time: 3.25 ms

```

```

[41]: %%time
      # Agregamos esas palabras a la lista de stopwords
      from collections import Counter

```

```

words = []
for anio in data_ln.columns:
    top = [word for (word, count) in top_dict[anio]]
    for t in top:
        words.append(t)

words

```

CPU times: user 591 µs, sys: 0 ns, total: 591 µs

Wall time: 605 µs

```

[41]: ['polideportivo',
      'lector',
      'cartas',
      'sellers',
      'best',
      'libro',
      'hoy',
      'dos',
      'aires',
      'buenos',
      'rural',
      'mejor',
      'xp',
      'controles',
      'día',
      'crece',
      'poder',
      'obras',
      'premiaron',
      'windows',
      'salto',
      'peso',
      'premio',
      'recibir',
      'país',
      'sorín',
      'exportador',
      'diaz',
      'pasa',
      'mundo',
      'lector',
      'cartas',
      'libro',
      'best',
      'cine',

```


'sellers',
'gran',
'nuevo',
'potter',
'harry',
'fiesta',
'mundo',
'arte',
'breves',
'año',
'vida',
'rock',
'música',
'hermano',
'nacion',
'polideportivo',
'hoy',
'argentino',
'viaje',
'crisis',
'semana',
'según',
'vuelve',
'muestra',
'empresas',
'lector',
'cartas',
'concurso',
'fiesta',
'argentina',
'bases',
'nuevo',
'mejor',
'novela',
'nueva',
'nacion',
'cine',
'música',
'gran',
'premio',
'polideportivo',
'ensayo',
'récord',
'ganó',
'potter',
'arte',
'año',

'noche',
'sudamericana',
'años',
'argentinos',
'libro',
'feedback',
'festival',
'publicidad',
'agenda',
'mejores',
'internet',
'videos',
'insólitos',
'argentina',
'lector',
'mundo',
'cartas',
'crisis',
'nueva',
'nuevo',
'mejor',
'argentinas',
'año',
'mundial',
'historia',
'primera',
'gran',
'bases',
'fiesta',
'gobierno',
'gritos',
'conocé',
'premio',
'arte',
'susurros',
'libro',
'vuelve',
'semana',
'mejores',
'videos',
'insólitos',
'internet',
'aires',
'buenos',
'cine',
'mundial',
'nuevos',

'gran',
'noche',
'año',
'argentina',
'lector',
'libro',
'argentinos',
'cómo',
'nuevo',
'feedback',
'llega',
'casa',
'mundo',
'polideportivo',
'dakar',
'cartas',
'nueva',
'consumo',
'según',
'días',
'récord',
'argentina',
'libro',
'gran',
'nuevo',
'cinco',
'polideportivo',
'hoy',
'historia',
'común',
'mejor',
'gadgets',
'cómo',
'geekmanía',
'lector',
'vuelve',
'ipad',
'año',
'encuentro',
'feria',
'noche',
'bafweek',
'cartas',
'mundo',
'digital',
'argentino',
'argentinas',

'mundial',
'rs',
'puro',
'tras',
'gran',
'nueva',
'argentina',
'fiesta',
'nuevo',
'festival',
'país',
'cómo',
'windows',
'mejor',
'mundo',
'noche',
'obra',
'gobierno',
'casa',
'libro',
'millones',
'lector',
'martín',
'grammy',
'industria',
'polideportivo',
'equipo',
'microsoft',
'moreno',
'premios',
'semana',
'argentino',
'puerto',
'feedback',
'cultural',
'agenda',
'argentina',
'libro',
'feria',
'mundo',
'nueva',
'buenos',
'papa',
'cómo',
'aires',
'nuevo',
'país',

'presentó',
'historia',
'cristina',
'vez',
'festival',
'regreso',
'pantallas',
'ciudad',
'fin',
'mejor',
'dos',
'gobierno',
'diez',
'francisco',
'apple',
'teatro',
'arte',
'mundial',
'nuevo',
'argentina',
'fiesta',
'hoy',
'cómo',
'mundo',
'arte',
'cine',
'aires',
'años',
'brasil',
'libro',
'buenos',
'ganó',
'año',
'boca',
'premio',
'días',
'oscar',
'nueva',
'hot',
'tres',
'cristina',
'tras',
'síntesis',
'rock',
'comienza',
'polideportivo',
'festival',

'nueva',
'fin',
'edición',
'argentina',
'mundo',
'año',
'fiesta',
'años',
'gran',
'mejor',
'lector',
'semana',
'síntesis',
'cartas',
'premios',
'the',
'final',
'dos',
'aires',
'buenos',
'argentino',
'feria',
'guía',
'futuro',
'nuevos',
'dakar',
'vuelve',
'noche',
'tras',
'libro',
'lector',
'cartas',
'verano',
'viene',
'nueva',
'fiesta',
'tan',
'muerte',
'dakar',
'negocio',
'foco',
'pudo',
'fin',
'extraño',
'fútbol',
'maratón',
'ces',

```
'segundo',  
'secretos',  
'arte',  
'ser',  
'argentina',  
'último',  
'terminó',  
'punta',  
'loeb',  
'boca',  
'lista',  
'libro',  
'tv']
```

```
[42]: Counter(words).most_common()
```

```
[42]: [('libro', 11),  
      ('lector', 9),  
      ('mundo', 9),  
      ('argentina', 9),  
      ('cartas', 8),  
      ('nuevo', 8),  
      ('nueva', 8),  
      ('polideportivo', 7),  
      ('mejor', 7),  
      ('gran', 7),  
      ('fiesta', 7),  
      ('año', 7),  
      ('arte', 6),  
      ('aires', 5),  
      ('buenos', 5),  
      ('noche', 5),  
      ('cómo', 5),  
      ('hoy', 4),  
      ('premio', 4),  
      ('cine', 4),  
      ('argentino', 4),  
      ('semana', 4),  
      ('vuelve', 4),  
      ('festival', 4),  
      ('mundial', 4),  
      ('dos', 3),  
      ('país', 3),  
      ('años', 3),  
      ('feedback', 3),  
      ('historia', 3),  
      ('gobierno', 3),
```

('dakar', 3),
('feria', 3),
('tras', 3),
('fin', 3),
('sellers', 2),
('best', 2),
('windows', 2),
('potter', 2),
('rock', 2),
('música', 2),
('nacion', 2),
('crisis', 2),
('según', 2),
('bases', 2),
('récord', 2),
('ganó', 2),
('argentinos', 2),
('agenda', 2),
('mejores', 2),
('internet', 2),
('videos', 2),
('insólitos', 2),
('argentinas', 2),
('nuevos', 2),
('casa', 2),
('días', 2),
('premios', 2),
('cristina', 2),
('boca', 2),
('síntesis', 2),
('rural', 1),
('xp', 1),
('controles', 1),
('día', 1),
('crece', 1),
('poder', 1),
('obras', 1),
('premiaron', 1),
('salto', 1),
('peso', 1),
('recibir', 1),
('sorín', 1),
('exportador', 1),
('diaz', 1),
('pasa', 1),
('harry', 1),
('breves', 1),

('vida', 1),
('hermano', 1),
('viaje', 1),
('muestra', 1),
('empresas', 1),
('concurso', 1),
('novela', 1),
('ensayo', 1),
('sudamericana', 1),
('publicidad', 1),
('primera', 1),
('gritos', 1),
('conocé', 1),
('susurros', 1),
('llega', 1),
('consumo', 1),
('cinco', 1),
('común', 1),
('gadgets', 1),
('geekmanía', 1),
('ipad', 1),
('encuentro', 1),
('bafweek', 1),
('digital', 1),
('rs', 1),
('puro', 1),
('obra', 1),
('millones', 1),
('martín', 1),
('grammy', 1),
('industria', 1),
('equipo', 1),
('microsoft', 1),
('moreno', 1),
('puerto', 1),
('cultural', 1),
('papa', 1),
('presentó', 1),
('vez', 1),
('regreso', 1),
('pantallas', 1),
('ciudad', 1),
('diez', 1),
('francisco', 1),
('apple', 1),
('teatro', 1),
('brasil', 1),

```
(('oscar', 1),
 ('hot', 1),
 ('tres', 1),
 ('comienza', 1),
 ('edición', 1),
 ('the', 1),
 ('final', 1),
 ('guía', 1),
 ('futuro', 1),
 ('verano', 1),
 ('viene', 1),
 ('tan', 1),
 ('muerte', 1),
 ('negocio', 1),
 ('foco', 1),
 ('pudo', 1),
 ('extraño', 1),
 ('fútbol', 1),
 ('maratón', 1),
 ('ces', 1),
 ('segundo', 1),
 ('secretos', 1),
 ('ser', 1),
 ('último', 1),
 ('terminó', 1),
 ('punta', 1),
 ('loeb', 1),
 ('lista', 1),
 ('tv', 1)])
```

```
[43]: %%time
# Las mas repetidas las descartaremos
add_stop_words = [word for word, count in Counter(words).most_common() if count <
→ 6]
add_stop_words
```

```
CPU times: user 3.35 ms, sys: 0 ns, total: 3.35 ms
Wall time: 2.2 ms
```

```
[43]: ['libro',
 'lector',
 'mundo',
 'argentina',
 'cartas',
 'nuevo',
 'nueva',
 'polideportivo',
```

```
'mejor',  
'gran',  
'fiesta',  
'año']
```

```
[44]: %%time  
# A partir de unos textos genericos, creamos un corpus linguistico (ejemplos de uso)  
# usamos un blog  
anios = ['2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013',  
        '2014', '2015', '2016']  
import nltk  
from nltk.corpus import PlaintextCorpusReader  
corpus_root = '/home/administrador/git/python/tp_ds/blog/'  
wordlists = PlaintextCorpusReader(corpus_root, '.*', encoding='latin-1')  
  
cfd = nltk.ConditionalFreqDist(  
    (word, genre)  
    for genre in anios  
    for w in wordlists.words(genre + '.txt')  
    for word in ['casa', 'mundo', 'tiempo', 'vida']  
    if w.lower().startswith(word) )  
cfd.plot()
```



CPU times: user 776 ms, sys: 152 μ s, total: 776 ms

Wall time: 773 ms

[44]: <matplotlib.axes._subplots.AxesSubplot at 0x7f230dc462d0>

```
[47]: %%time
# Como no alcanzan las palabras del diccionario, debemos agregar mas stopwords
# extraidas manualmente del texto
from sklearn.feature_extraction import text
from sklearn.feature_extraction.text import CountVectorizer

# Leemos los textos limpios
data_ln_clean = pd.read_pickle('data_ln_clean.pkl')

# Agregamos algunas stopwords de nuestra cosecha
with open('spanish.txt') as f:
    stop_words = f.read().splitlines()
for pal in add_stop_words:
    stop_words.append(pal)
more_stop_words=['alex','andr s','asi','andres','as ','a o','alejandro','alfonso','all ','algu
                'basdala','bernardo','bien',
```

```

        ↪ 'cosa', 'cosas', 'costoya', 'costa', 'cinco', 'celoni', 'cuatro', 'cómo', 'casi', 'colo', 'caprio', 'č
        'dos', 'dice', 'decir', 'días', 'dije', 'digo', 'diez',
        'ésa', 'ésas', 'ése', 'ésos', 'ésta', 'ésta', 'éste', 'ésto',
        ↪ 'éstos',
        'fernando', 'fenwick',
        'gelós', 'gente',
        'hornby', 'hoy', 'horacio', 'horas', 'hará', 'hans', 'hacia', 'haber',
        'iveta',
        'jesús', 'jorge', 'juan',
        'karen',
        'lucas', 'luego', 'luis',
        ↪
        ↪ 'mirta', 'mientras', 'menos', 'mónica', 'medio', 'mil', 'moncho', 'momento', 'mañana', 'mejor',
        'narcís', 'número', 'noche', 'nadie',
        'ojos',
        'primer', 'primera', 'pase', 'pablo', 'pepe', 'pack', 'peter',
        ↪ 'pues', 'prieto', 'politto', 'pol', 'paola', 'puede', 'próximo', 'podrán', 'podía',
        'quizá', 'quizás', 'quince', 'quién', 'quiero',
        'rato',
        ↪
        ↪ 'sólo', 'solamente', 'sakhan', 'šeredova', 'seis', 'šeredová', 'seselovsky', 'solo', 'salas', 'sant'
        'tres', 'tan', 'todas', 'trece', 'toda', 'todavía', 'tarde', 'tener',
        'uno', 'usted',
        'veces', 'ver', 've', 'vos', 'va', 'voy',
        'waiser', 'woung'
    ]
for pal in more_stop_words:
    stop_words.append(pal)

cv = CountVectorizer(stop_words=stop_words)
data_ln_cv = cv.fit_transform(data_ln_clean.Texto)
data_ln_stop = pd.DataFrame(data_ln_cv.toarray(), columns=cv.
    ↪get_feature_names())
data_ln_stop.index = data_ln_clean.index

# guardamos por las dudas
import pickle
pickle.dump(cv, open("cv_stop.pkl", "wb"))
data_ln_stop.to_pickle("dtm_stop.pkl")

```

CPU times: user 69.7 ms, sys: 0 ns, total: 69.7 ms
Wall time: 69.5 ms

```
[48]: %%time
# Terminal / Anaconda Prompt: conda install -c conda-forge wordcloud
from wordcloud import WordCloud

wc = WordCloud(stopwords=stop_words, background_color="white", colormap="Dark2",
               max_font_size=150, random_state=42)
```

CPU times: user 226 µs, sys: 0 ns, total: 226 µs

Wall time: 247 µs

```
[50]: %%time
# Dibujamos las nubes de palabras
import matplotlib.pyplot as plt

plt.rcParams['figure.figsize'] = [16,12]

#anios =
→ ['2004', '2005', '2006', '2007', '2008', '2009', '2010', '2011', '2012', '2013', '2014', '2015']

# Create subplots for each anio
for index, anio in enumerate(data_ln.columns):
    wc.generate(data_ln_clean.Texto[anio])
    plt.subplot(4, 3, index+1)
    plt.imshow(wc, interpolation="bilinear")
    plt.axis("off")
    plt.title(anios[index])

plt.show()
```



CPU times: user 2.49 s, sys: 128 ms, total: 2.62 s
Wall time: 2.65 s

6 5 - Análisis de Sentimiento

```
[51]: %%time
# Leeremos el corpus que aún preserva el orden de las palabras
import pandas as pd

data_ln = pd.read_pickle('corpus_ln.pkl')
data_ln
```

CPU times: user 6.81 ms, sys: 42 µs, total: 6.85 ms
Wall time: 4.54 ms

```
[51]:
```

2006 bajo el signo del terror apple renovó toda la familia ipod por las canchas argentinos en francia ;un loco suelto en hollywood cartas de lector...

2007 cómo saber si tu computadora será capaz de usar windows vista sin inconvenientes juegos que acaban de llegar y algunos próximos estrenos la

impor...

2008 cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda la música y los celulares el futuro de nuestra lengua bases del concur...

2009 la mejor publicidad del año episodio ii algo sobre los vascos mundo extremo un recorrido por los tesoros culturales de la argentina el largo cami...

2010 el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...

2011 dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válid...

2012 concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...

2013 agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay escrit...

2014 el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...

2015 sospechan que a lola chomnalez la habrían asesinado el sábado en más del de las ventas y pagos online se hará desde un móvil versos peregrinos...

2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de medalla tiene la argentina en río la última etapa de las pelucas qué podem...

	full_name
2006	2006
2007	2007
2008	2008
2009	2009
2010	2010
2011	2011
2012	2012
2013	2013
2014	2014
2015	2015
2016	2016

```
[52]: %%time
# Creamos funciones auxiliares para procesar el texto
# Terminal / Anaconda Navigator: conda install -c conda-forge textblob
from textblob import TextBlob

pol = lambda x: TextBlob(x).sentiment.polarity
pol2 = lambda x: x.sentiment.polarity
sub = lambda x: TextBlob(x).sentiment.subjectivity
sub2 = lambda x: x.sentiment.subjectivity

# Realmente lo traducimos al inglés pues el analisis de sentimiento de TextBlob
→no funciona en Español :(
```



```

traducir = lambda x: TextBlob(x).translate(to="en")

data_ln['blob_en'] = data_ln['Texto'].apply(traducir)
data_ln['polarity'] = data_ln['blob_en'].apply(pol2)
data_ln['subjectivity'] = data_ln['blob_en'].apply(sub2)
data_ln

```

CPU times: user 1.01 s, sys: 1.26 ms, total: 1.01 s
Wall time: 4.06 s

[52]:

	Texto \
2006	bajo el signo del terror apple renovó toda la familia ipod por las canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007	cómo saber si tu computadora será capaz de usar windows vista sin inconvenientes juegos que acaban de llegar y algunos próximos estrenos la impor...
2008	cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda la música y los celulares el futuro de nuestra lengua bases del concur...
2009	la mejor publicidad del año episodio ii algo sobre los vascos mundo extremo un recorrido por los tesoros culturales de la argentina el largo cami...
2010	el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...
2011	dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válid...
2012	concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...
2013	agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay escrit...
2014	el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...
2015	sospechan que a lola chomnalez la habrían asesinado el sábado en más del de las ventas y pagos online se hará desde un móvil versos peregrinos...
2016	pinamar una nueva edición de la maratón del desierto ¿qué chances de medalla tiene la argentina en río la última etapa de las pelucas qué podem...

	full_name \
2006	2006
2007	2007
2008	2008
2009	2009
2010	2010
2011	2011
2012	2012
2013	2013
2014	2014

2015 2015
2016 2016

```

blob_en \
2006 (U, n, d, e, r, , t, h, e, , s, i, g, n, , o, f, , t, e, r, r, o, r,
,, , A, p, p, l, e, , r, e, n, e, w, e, d, , t, h, e, , e, n, t, i, r...
2007 (h, o, w, , t, o, , k, n, o, w, , i, f, , y, o, u, r, , c, o, m, p,
u, t, e, r, , w, i, l, l, , b, e, , a, b, l, e, , t, o, , u, s, e, ...
2008 (r, e, a, d, e, r, s, ', , l, e, t, t, e, r, s, , t, h, e, , m, y, s,
t, e, r, y, , o, f, , r, i, m, b, a, u, d, , o, n, , t, h, e, , b, e...
2009 (t, h, e, , b, e, s, t, , p, u, b, l, i, c, i, t, y, , o, f, , t, h,
e, , y, e, a, r, , e, p, i, s, o, d, e, , i, i, , s, o, m, e, t, h, i...
2010 (t, h, e, , b, e, g, i, n, n, i, n, g, , o, f, , t, h, e, , d, a, k,
a, r, , w, a, s, , a, , p, u, r, e, , e, m, o, t, i, o, n, , i, n, ...
2011 (d, a, k, a, r, , m, e, n, d, o, z, a, , b, r, i, n, g, s, , p, a, r,
t, i, c, i, p, a, n, t, s, , u, n, d, e, r, , t, h, e, , h, e, a, t, ...
2012 (c, o, n, t, e, s, t, , t, h, e, , a, l, b, u, m, , o, f, , A, r, g,
e, n, t, i, n, a, , b, a, s, e, s, , v, i, c, t, i, m, s, , w, r, o, n...
2013 (t, h, e, y, , a, d, d, , a, , g, i, r, l, , t, h, e, y, , f, i, l,
m, , t, h, e, , b, e, a, t, i, n, g, , a, n, d, , t, h, e, y, , u, p...
2014 (T, h, e, , b, i, g, , b, o, y, , w, h, o, , c, o, n, t, i, n, u, e,
s, , t, o, , s, u, r, p, r, i, s, e, , u, s, , o, n, , t, i, p, , t...
2015 (t, h, e, y, , s, u, s, p, e, c, t, , t, h, a, t, , l, o, l, a, , c,
h, o, m, n, a, l, e, z, , w, o, u, l, d, , h, a, v, e, , k, i, l, l, e...
2016 (p, i, n, n, a, m, a, r, , a, , n, e, w, , e, d, i, t, i, o, n, , o,
f, , t, h, e, , m, a, r, a, t, h, o, n, , o, f, , t, h, e, , d, e, s...

```

	polarity	subjectivity
2006	0.203939	0.394890
2007	0.208520	0.431694
2008	0.154556	0.424543
2009	0.243064	0.465497
2010	0.212595	0.440779
2011	0.137775	0.433609
2012	0.182623	0.437851
2013	0.176689	0.433935
2014	0.146375	0.420615
2015	0.145093	0.414973
2016	0.099329	0.425792

```

[54]: %%time
# Dibujemos los resultados
import matplotlib.pyplot as plt

plt.rcParams['figure.figsize'] = [10, 8]

for index, anio in enumerate(data_ln.index):

```

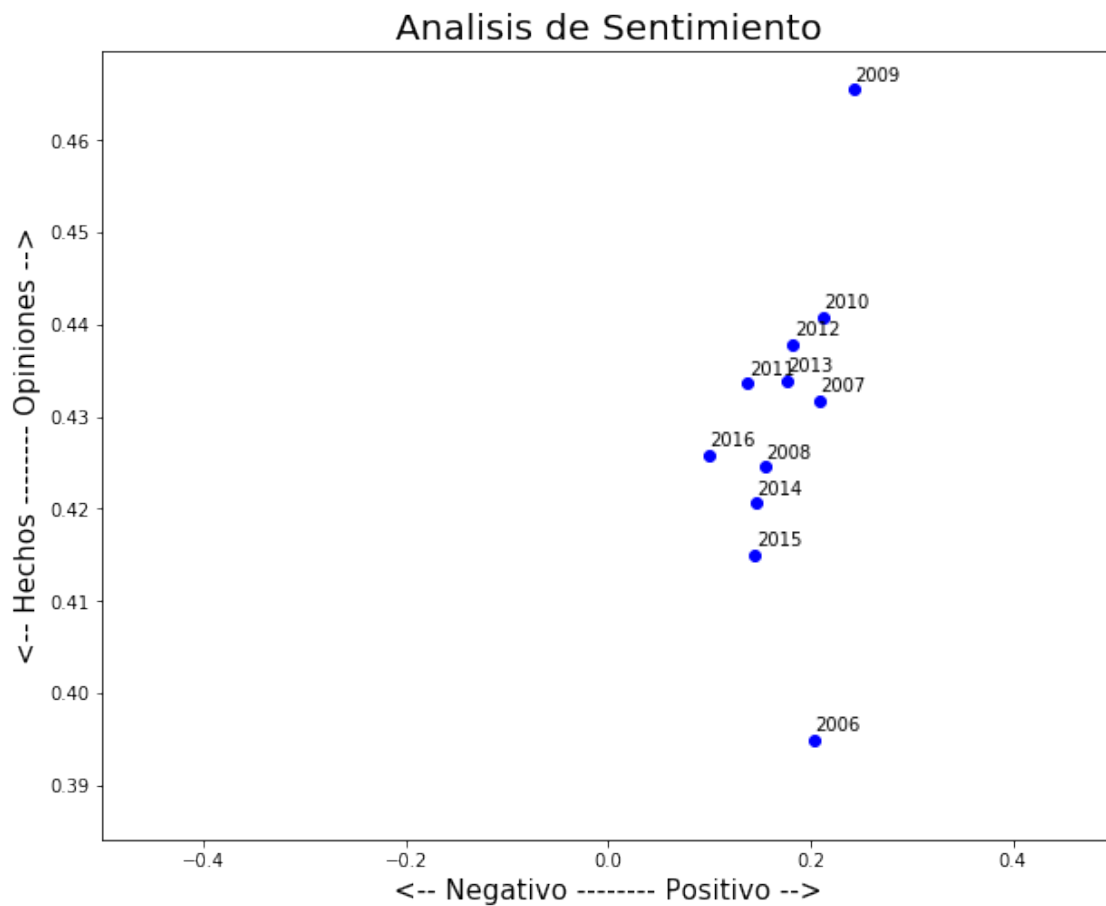
```

x = data_ln.polarity.loc[anio]
y = data_ln.subjectivity.loc[anio]
plt.scatter(x, y, color='blue')
#plt.text(x+.001, y+.001, data['full_names'][index], fontsize=10)
plt.text(x+.001, y+.001, anio, fontsize=10)
#plt.xlim(-0.051, 0.152)
plt.xlim(-0.5, 0.5)

plt.title('Análisis de Sentimiento', fontsize=20)
plt.xlabel('<-- Negativo ----- Positivo -->', fontsize=15)
plt.ylabel('<-- Hechos ----- Opiniones -->', fontsize=15)

plt.show()

```



CPU times: user 213 ms, sys: 11.6 ms, total: 225 ms
Wall time: 223 ms

6.1 Variacion del sentimiento con el tiempo

```
[55]: %%time
      # Partimos cada año en 12 piezas correspondiente a los meses
      import numpy as np
      import math

      def split_text(text, n=12):

          length = len(text)
          size = math.floor(length / n)
          start = np.arange(0, length, size)

          split_list = []
          for piece in range(n):
              split_list.append(text[start[piece]:start[piece]+size])
          return split_list
```

CPU times: user 31 μ s, sys: 2 μ s, total: 33 μ s
Wall time: 43.9 μ s

```
[56]: # Veamos que datos tenemos
      data_ln
```

```
[56]:
```

	Texto \
2006	bajo el signo del terror apple renovó toda la familia ipod por las canchas argentinos en francia ¿un loco suelto en hollywood cartas de lector...
2007	cómo saber si tu computadora será capaz de usar windows vista sin inconvenientes juegos que acaban de llegar y algunos próximos estrenos la impor...
2008	cartas de lectores el misterio de rimbaud en la playa al ritmo de la moda la música y los celulares el futuro de nuestra lengua bases del concur...
2009	la mejor publicidad del año episodio ii algo sobre los vascos mundo extremo un recorrido por los tesoros culturales de la argentina el largo cami...
2010	el inicio del dakar fue a pura emoción en buenos aires un año para consolidar el crecimiento parte de los festejos por el bicentenario gritos y su...
2011	dakar mendoza aporta participantes bajo el calor de la admiración cómo manejarse con la nueva ortografía ayer fue censurado pero hoy es válid...
2012	concurso el album de la argentina bases víctimas erradas en la masacre de rosario la inflación hace bajar el ingreso de turistas extranjeros la l...
2013	agreden a una chica filman la paliza y suben el video a las redes sociales actualidad críticas puerta abierta a la escuela del futuro hay escrit...
2014	el niño grande que nos sigue sorprendiendo en punta la fiesta siguió en la playa nervo mujeres bellas y de beats fuertes navidad junto al lago...
2015	sospechan que a lola chomnalez la habrían asesinado el sábado en más del

de las ventas y pagos online se hará desde un móvil versos peregrinos...
 2016 pinamar una nueva edición de la maratón del desierto ¿qué chances de
 medalla tiene la argentina en río la última etapa de las pelucas qué podem...

	full_name \	blob_en \
2006	2006	
2007	2007	
2008	2008	
2009	2009	
2010	2010	
2011	2011	
2012	2012	
2013	2013	
2014	2014	
2015	2015	
2016	2016	
2006	(U, n, d, e, r, , t, h, e, , s, i, g, n, , o, f, , t, e, r, r, o, r, , , A, p, p, l, e, , r, e, n, e, w, e, d, , t, h, e, , e, n, t, i, r...	
2007	(h, o, w, , t, o, , k, n, o, w, , i, f, , y, o, u, r, , c, o, m, p, u, t, e, r, , w, i, l, l, , b, e, , a, b, l, e, , t, o, , u, s, e, ...	
2008	(r, e, a, d, e, r, s, ', , l, e, t, t, e, r, s, , t, h, e, , m, y, s, t, e, r, y, , o, f, , r, i, m, b, a, u, d, , o, n, , t, h, e, , b, e...	
2009	(t, h, e, , b, e, s, t, , p, u, b, l, i, c, i, t, y, , o, f, , t, h, e, , y, e, a, r, , e, p, i, s, o, d, e, , i, i, , s, o, m, e, t, h, i...	
2010	(t, h, e, , b, e, g, i, n, n, i, n, g, , o, f, , t, h, e, , d, a, k, a, r, , w, a, s, , a, , p, u, r, e, , e, m, o, t, i, o, n, , i, n, ...	
2011	(d, a, k, a, r, , m, e, n, d, o, z, a, , b, r, i, n, g, s, , p, a, r, t, i, c, i, p, a, n, t, s, , u, n, d, e, r, , t, h, e, , h, e, a, t, ...	
2012	(c, o, n, t, e, s, t, , t, h, e, , a, l, b, u, m, , o, f, , A, r, g, e, n, t, i, n, a, , b, a, s, e, s, , v, i, c, t, i, m, s, , w, r, o, n...	
2013	(t, h, e, y, , a, d, d, , a, , g, i, r, l, , t, h, e, y, , f, i, l, m, , t, h, e, , b, e, a, t, i, n, g, , a, n, d, , t, h, e, y, , u, p...	
2014	(T, h, e, , b, i, g, , b, o, y, , w, h, o, , c, o, n, t, i, n, u, e, s, , t, o, , s, u, r, p, r, i, s, e, , u, s, , o, n, , t, i, p, , t...	
2015	(t, h, e, y, , s, u, s, p, e, c, t, , t, h, a, t, , l, o, l, a, , c, h, o, m, n, a, l, e, z, , w, o, u, l, d, , h, a, v, e, , k, i, l, l, e...	
2016	(p, i, n, n, a, m, a, r, , a, , n, e, w, , e, d, i, t, i, o, n, , o, f, , t, h, e, , m, a, r, a, t, h, o, n, , o, f, , t, h, e, , d, e, s...	

	polarity	subjectivity
2006	0.203939	0.394890
2007	0.208520	0.431694
2008	0.154556	0.424543
2009	0.243064	0.465497
2010	0.212595	0.440779

2011	0.137775	0.433609
2012	0.182623	0.437851
2013	0.176689	0.433935
2014	0.146375	0.420615
2015	0.145093	0.414973
2016	0.099329	0.425792

```
[57]: %%time
# Juntamos todos los textos
list_pieces = []
for t in data_ln.blob_en:#transcript:
    split = split_text(t,12)
    list_pieces.append(split)
#list_pieces
```

CPU times: user 14.3 ms, sys: 108 µs, total: 14.4 ms
Wall time: 13 ms

```
[58]: # Cada lista tiene n elementos
len(list_pieces)
```

[58]: 11

```
[59]: # Ahora lo partimos en piezas de 10
len(list_pieces[0])
```

[59]: 12

```
[60]: %%time
# Calculamos la polaridad de cada pieza de texto

polarity_transcript = []
for lp in list_pieces:
    polarity_piece = []
    for p in lp:
        #polarity_piece.append(TextBlob(p).translate(to="en").sentiment.
        ↪polarity)
        polarity_piece.append(p.sentiment.polarity)
    polarity_transcript.append(polarity_piece)

polarity_transcript
```

CPU times: user 141 ms, sys: 263 µs, total: 141 ms
Wall time: 141 ms

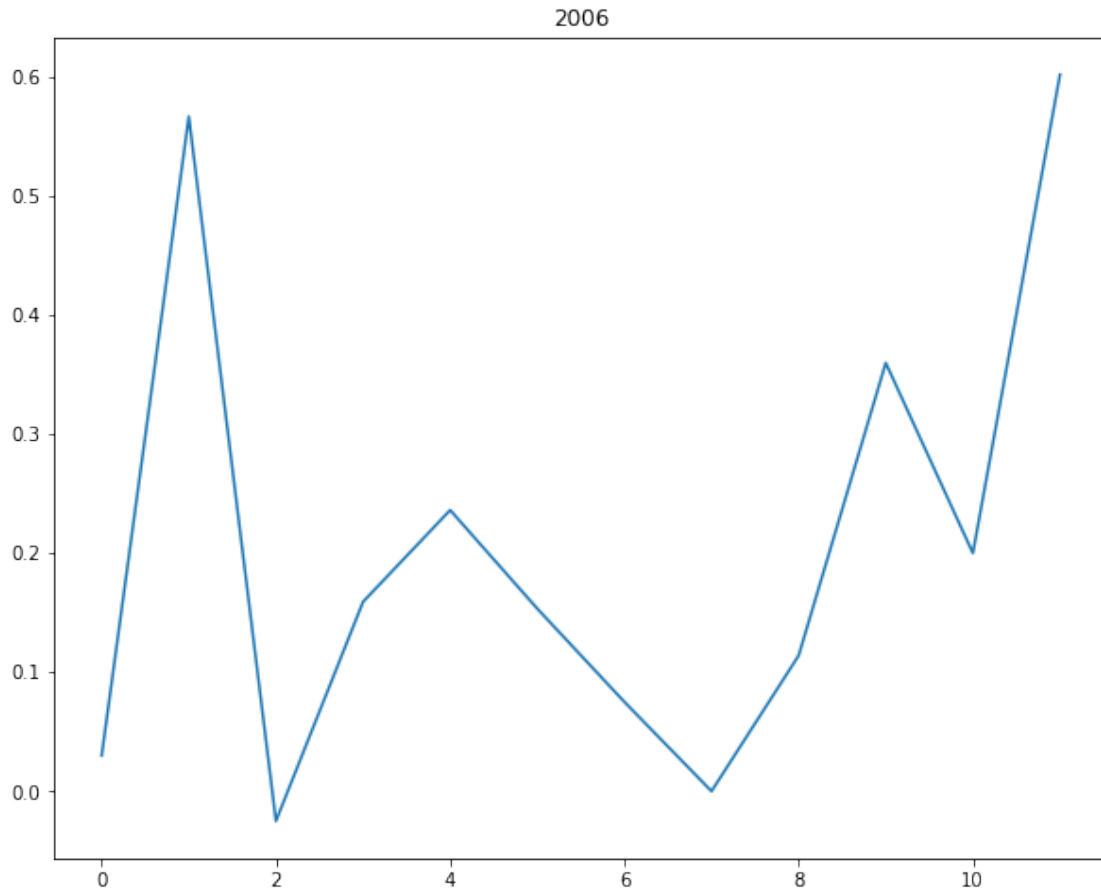
```
[60]: [[0.030027472527472527,
0.5666666666666667,
```

-0.025,
0.1590909090909091,
0.23611111111111113,
0.1533333333333332,
0.07500000000000001,
0.0,
0.11393939393939392,
0.35977272727272724,
0.2,
0.6018939393939394],
[0.18743506493506498,
0.2262053571428571,
0.3377651515151516,
0.24568181818181825,
0.2696428571428572,
0.07091750841750841,
-0.016666666666666663,
0.2872194163860831,
0.09769904009034444,
0.24350649350649348,
0.30862193362193363,
0.13506198347107434],
[0.24696969696969698,
0.018805045529183464,
0.08669696969696972,
0.02100550964187328,
0.25555555555555554,
0.0709486166007905,
0.1046422558922559,
0.1457260101010101,
0.19071146245059287,
0.22352430555555555,
0.18614610389610386,
0.2804343434343435],
[0.302962037962038,
0.25563725490196076,
0.18233333333333335,
0.2135776723276723,
0.27654428904428907,
0.25736531986531985,
0.2348413825757576,
0.1604511019283747,
0.2112889983579639,
0.25669191919191925,
0.25415669856459333,
0.2866666666666667],
[0.19978634268956852,

0.30406336088154273,
0.15015151515151515,
0.23346680201518913,
0.2176548089591568,
0.19743991640543362,
0.15281713236258693,
0.271900826446281,
0.167283950617284,
0.26518065268065266,
0.2169372294372294,
0.17464285714285713],
[0.02269570707070707,
0.2619047619047619,
0.12202797202797201,
0.008080808080808084,
0.12460127591706535,
0.05016835016835016,
0.27370129870129867,
0.14602415128730917,
0.13218923933209648,
0.047544409613375124,
0.2412146676852559,
0.29027777777777775],
[0.0891919191919192,
0.2845423936333027,
0.1358180014430014,
0.19861111111111113,
0.257027972027972,
0.19845454545454544,
0.1278409090909091,
0.18129058441558443,
0.1785511363636364,
0.13698019152564608,
0.2194638694638694,
0.0778030303030303],
[0.1240952380952381,
0.32306818181818187,
0.217151934087418,
0.19495109828443163,
0.2152597402597402,
0.08115530303030304,
0.13288770053475937,
0.11359427609427608,
0.11006060606060605,
0.1779505582137161,
0.3172219085262563,
0.09175407925407926],


```
[0.12148268398268396,
 0.15652442795299937,
 -0.0244077134986226,
 0.18622668997668998,
 0.19427609427609424,
 0.0975691699604743,
 -0.033525733525733514,
 0.33744019138755976,
 0.1830650252525253,
 0.15889125242573518,
 0.13626373626373628,
 0.25747474747474747],
[0.18068673750491931,
 0.2681439393939394,
 0.10254820936639118,
 0.18783670033670036,
 0.17129766067266067,
 0.2970238095238095,
 0.023940796857463524,
 0.03668132942326491,
 0.14553796897546897,
 0.08095238095238096,
 0.2516534391534392,
 0.0628902662993572],
[0.06818181818181818,
 0.1477272727272727,
 0.11212121212121212,
 0.016666666666666673,
 0.028333333333333332,
 0.0,
 0.12499999999999999,
 -0.1625,
 0.4166666666666667,
 0.16666666666666666,
 0.0,
 0.0028409090909090884]]
```

```
[61]: %%time
# Mostramos solo el etalle del primer año
plt.plot(polarity_transcript[0])
plt.title(data_ln['full_name'].index[0])
plt.show()
```

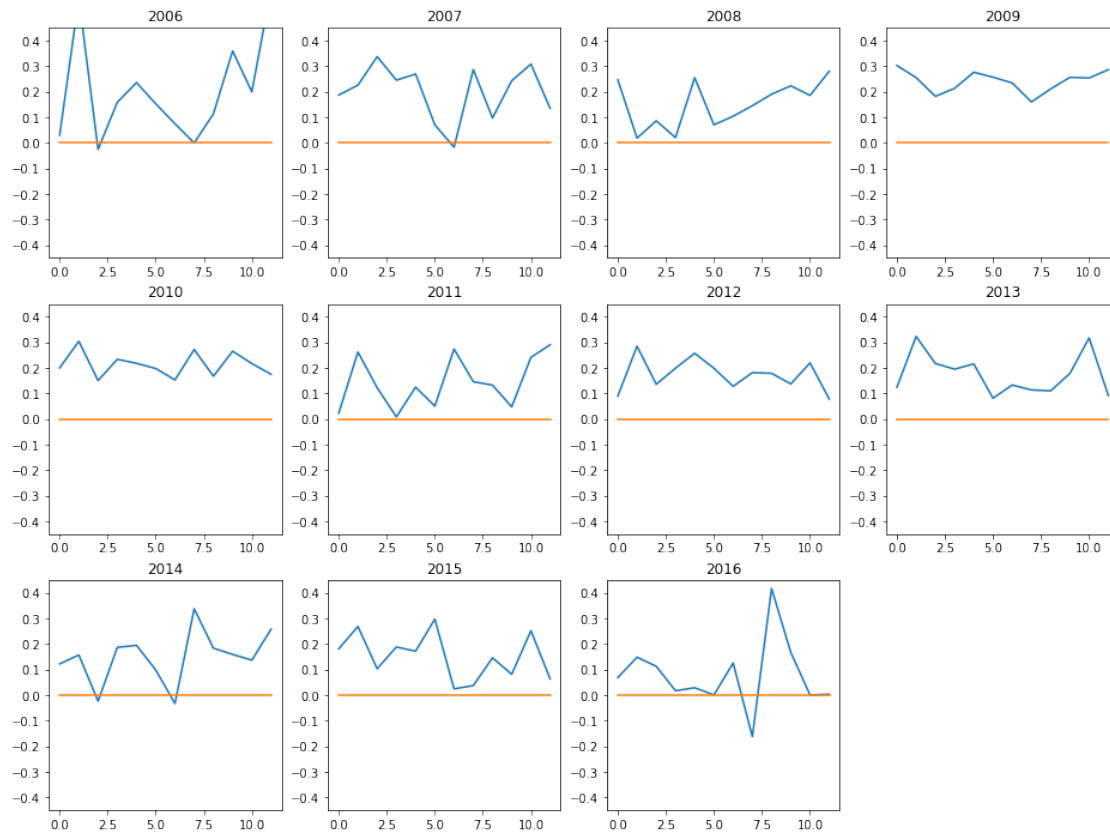


CPU times: user 173 ms, sys: 251 μ s, total: 173 ms
 Wall time: 170 ms

```
[62]: %%time
      # Hacemos todos los años
      plt.rcParams['figure.figsize'] = [16, 12]

      for index, anio in enumerate(data_ln.index):
          plt.subplot(3, 4, index+1)
          plt.plot(polarity_transcript[index])
          plt.plot(np.arange(0,12), np.zeros(12))
          #plt.title(data['full_name'][index])
          plt.title(anio)
          plt.ylim(ymin=-.45, ymax=.45)

      plt.show()
```



CPU times: user 1.1 s, sys: 7.91 ms, total: 1.11 s
Wall time: 1.1 s