Elemental analysis of the Spanish Government Twitter Account in 2020 with Mathematica Software (Part I)

Jesús Martín Ovejero.

Abstract: In the last ten years, programming languages as Python or R, have been established as indispensable tools in the armory of every Data Scientist, but, the new features of Wolfram Mathematica together with the fast development of other Wolfram Technologies allow us to suggest that Python and R hegemony is nearing its end. In the present notebook, we do an elemental analysis of the Spanish Government Twitter account using the software Mathematica. This short essay aims to show the potential use of Wolfram Technologies to analyze and manipulate data. This is the first part of a series of notebooks to show the versatility of Wolfram Language to analyze data and its potential use in finance analysis.

This notebook will be focused on the following points:

1) Extract information of a user Twitter account and bypass the rate limit of the API.

2) Use basic visualization tools to explore the data we have obtained.

3) Clean the data and perform a sentimental analysis.

To download the notebook go to: https://github.com/JMartinOvejero/Data-Analysis-with-Wolfram-Mathematica

I would like to thank Wolfram U for the opportunity to learn from so many experts in the field.

In the first place, to collect the tweets of the Spanish Government, we should connect with Twitter. There is a built-in API that allow us to access to Twitter:

```
In[ ]:= twitter = ServiceConnect["Twitter", SaveConnection → True]
                  conecta servicio          guarda conexión   verdade
```

```
Out[ ]= ServiceObject[ Twitter
                       Not Connected ]
```

The Spanish Government nickname at Twitter is @desdelamoncloa. Let us download the last 3222 tweets (rate limiting of Twitter API) posted by the Spanish Government Twitter account and let us store them in a variable called SpainTweets:

```
In[ ]:= SpainTweets = twitter["TweetList",
          "Username" → "desdelamoncloa", "MaxItems" → 5000, "Elements" → "FullData"];
                                         número máximo de items
```

```
In[ ]:= Dimensions[SpainTweets]
        dimensiones
```

```
Out[ ]= {3222, 11}
```

To see the attributes of the data we have collected, let us consider a random Tweet:

```
In[ ]:= RandomSample[SpainTweets, 1];
        muestra aleatoria
        %[[1]]
```

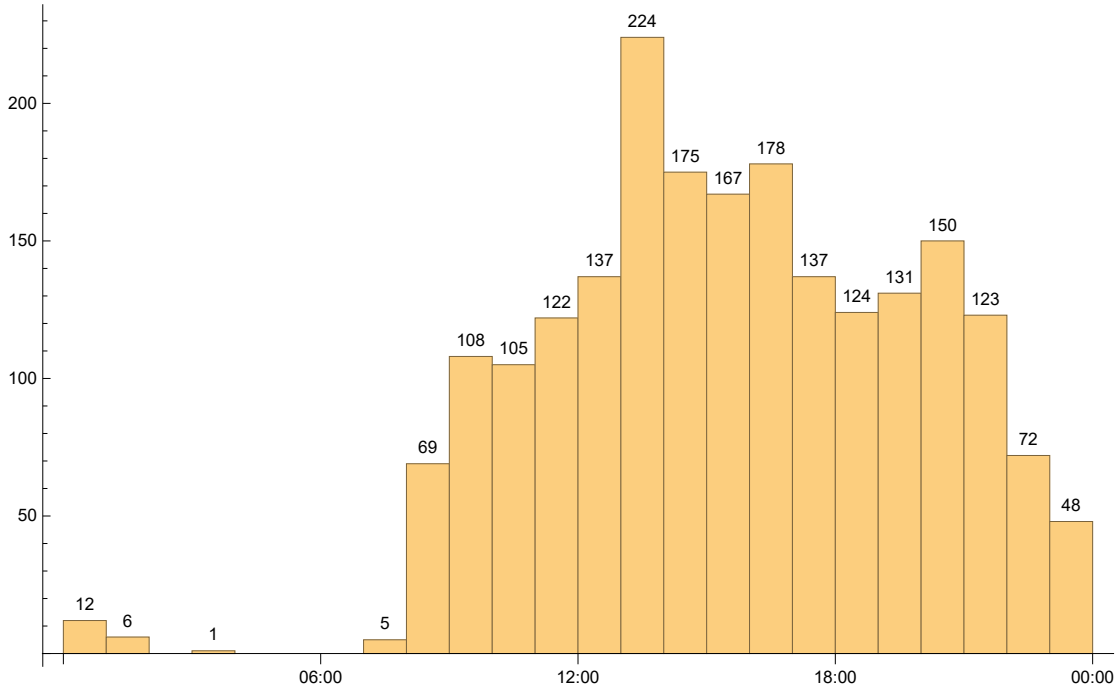
Out[ ]=

ID	1 221 765 726 501 056 515
Text	En el 75 aniversario de la liberación de Auschwitz, el @Senadoesp acoge el ac
Date	TimeZoneConvert[DateObject[{"2020 Jan 27 12:03:43", {"Year", "
Hashtags	{Holocausto, Auschwitz75, HolocaustMemorialDay}
Location	—
Username	desdelamoncloa
Name	La Moncloa
ProfileImageThumbnail	https://pbs.twimg.com/profile_images/1269184892363...
RetweetCount	44
FavoriteCount	71
URL	https://twitter.com/desdelamoncloa/status/12217657...

With two lines of code, and using the internal architecture of Mathematica, we can visualize the number of tweets posted on each day of the week and also at what time were posted them:

In[•]:=
```
DateHistogram[SpainTweets[All, "Date"], "Hour",
 DateReduction → "Day", LabelingFunction → Above, ImageSize → Large]
```
histograma de fechas · todo · fecha · reducción de fechas · función de etiquetado · encima · tamaño de i··· · grande

Out[•]=



Let us obtain the most popular tweets posted by @desdelamoncloa, i.e the Tweets with higher Favorite count and Retweet count.

In[•]:=
```
RetCount = Floor[SpainTweets[Max, "RetweetCount"]]
```
entero inferior · máximo

Out[•]=
```
11 750
```

In[•]:=
```
FavCount = Floor[SpainTweets[Max, "FavoriteCount"]]
```
entero inferior · máximo

Out[•]=
```
2497
```

In[•]:=
```
SpainTweets[
  Select[
   #RetweetCount ≥ RetCount ||
    #FavoriteCount ≥ FavCount &],
 {"Text", "Hashtags", "FavoriteCount", "RetweetCount"}]
```
selecciona · texto

Out[•]=

Text	Hasht
#EsteVirusLoParamosUnidos ⋱	{Este\
RT @SaludPublicaEs: Fernando Simón, de @sanidadgob, explica por qué es important ⋱	{}

To perform an analysis of the Hashtags used by the Twitter account, we can sort the Hashtags by

the number of times they were posted:

In[ ]:=
```
SpainHashtags = SpainTweets[Flatten /* Counts /* ReverseSort, "Hashtags"]
                            aplana   conteos   ordena al revés
```

Out[ ]=

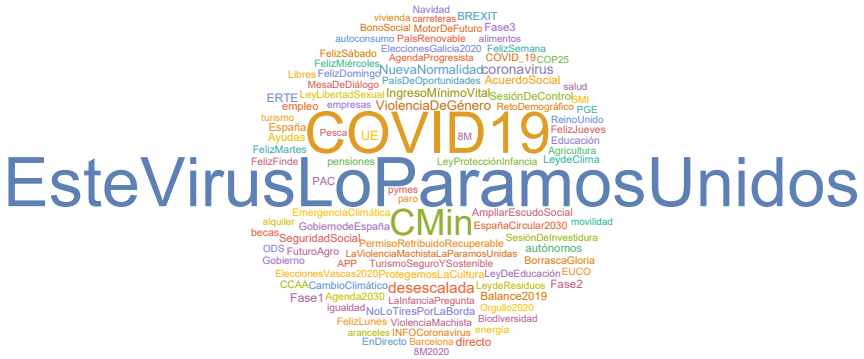
EsteVirusLoParamosUnidos	757
COVID19	612
CMin	335
desescalada	61
NuevaNormalidad	47
ViolenciaDeGénero	46
UE	45
coronavirus	43
Fase1	36
IngresoMínimoVital	36
directo	32
ERTE	31
AcuerdoSocial	30
Fase2	27
autónomos	27
Balance2019	26
Fase3	23
SeguridadSocial	23
PAC	22
SesiónDeControl	22

K < showing 1–20 of 858 > K

To visualize the frequency of the Hashtags we can use a word cloud:

In[ ]:=
```
SpainTweets[Flatten /* WordCloud, "Hashtags"]
            aplana   nube de palabras
```

Out[ ]=



Using the same techniques, we can create a word cloud out of the text content in the tweets in order to infer the main topics discussed in the tweets. Before doing this operation we must clean

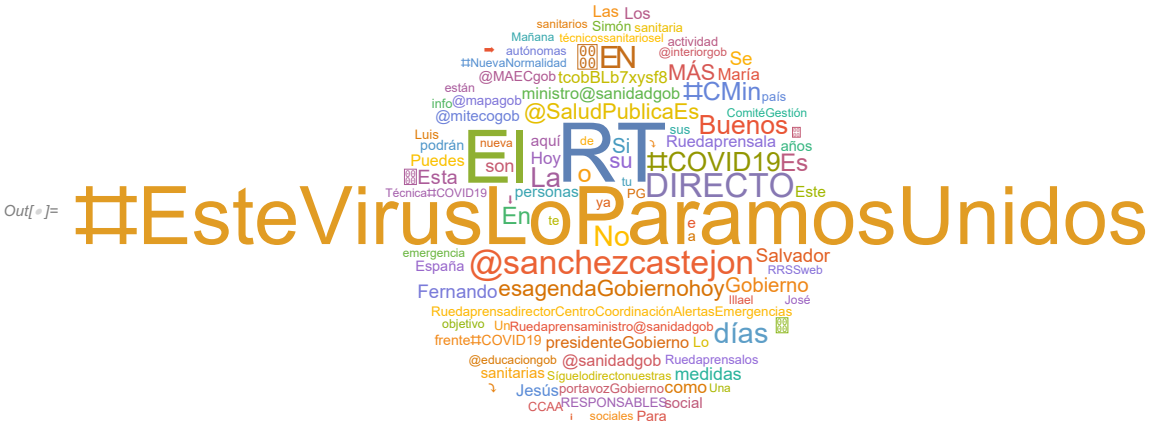
our data. With this aim, we remove undesired words and characters by defining a function called CleanText. In order to construct CleanText we must define a function to remove stopwords in Spanish.

```
stopwords = {" un ", " una ", " uno ", " unos ", " unas ", " nosotros ", " nosotras ",
    " vosotros ", " vosotras ", " el ", " El ", " la ", " La ", " han ", " lo ",
    " los ", " que ", " y ", " a ", " ante ", " bajo ", " con ", " contra" , " de ",
    " desde ", " en ", " entre ", " hacia ", " hasta ", " para ", " Para ", " Este ",
    " Esta ", " La ", " las ", " se" , " por ", " según ", " sobre ", " tras ",
    " también ", " por ", " ser ", " somos ", " esta ", " está ", " estamos ",
    " somos ", " estais ", " hacer ", " haciendo " , " otro ", " algún ", " porque ",
    " por qué ", " sin ", " usar ", " ciertos ", " cierto ", " todo ", " todos ",
    " todas ", " del ", " al ", " No ", " Si "; " más ", " ha ", " son " };

CleanText[text_String] := StringDelete[
                          elimina cadena de caracteres
  StringDelete[text, stopwords], {Except[{"@", "#", "_"}, PunctuationCharacter],
  elimina cadena de caracteres          excepto          caracter de puntuación
   "http" ~~ Except[WhitespaceCharacter]}]
             excepto  caracter de espacio en blanco
```

```
In[ ]:= SpainTweets[All /* StringSplit /* WordCloud, "Text", CleanText]
        todo  subdivide cadenas···  nube de palab···  texto
```

Out[ ]=



We can get a list of the mentioned users, but before this, we must write a function to extract user mentions and add them back to the structured data:

```
In[ ]:= getUserMentions[tweet_Association] :=
  Module[{modifiedTweet = tweet}, AssociateTo[modifiedTweet, "Usermentions" → Flatten[
  módulo                                  asocia a                                aplana
      StringCases[tweet["Text"], "@" ~~ u : (LetterCharacter | DigitCharacter | "_") .. ~~
      casos de cadena de cara··  texto      letra                dígito como caracter
          (WhitespaceCharacter | PunctuationCharacter | EndOfString) :→ u]]]];
          caracter de espacio en bla···  caracter de puntuación  final de cadena de caracteres

In[ ]:= SpainTweets = Dataset[getUserMentions /@ Normal @SpainTweets];
       conjunto de datos                normal
```

```
In[•]:= MentionUsers = Normal[SpainTweets[Flatten /* DeleteDuplicates, "Usermentions"]]
                       normal          aplana   elimina repeticiones
```

```
Out[•]= {policia, guardiacivil, glblctzn, sanchezcastejon, Congreso_Es, GAFSPfund, GlblCtzn,
 VSocialGob, INCIBE, mitecogob, MAECgob, mapagob, sanidadgob, Haciendagob,
 _minecogob, ICOgob, Defensagob, paradores, empleogob, inclusiongob, Yolanda_Diaz_,
 joseluisescriva, Senadoesp, gavi, UniversidadGob, salvadorilla, NadiaCalvino,
 ComisionEuropea, boegob, UE_Comision, educaciongob, territorialgob, Alimentacion_es,
 sanidad, igualdad, CasaReal, interiorgob, oapngob, mincoturgob, justiciagob,
 EU_Comm, CooperacionESP, proteccioncivil, SaludPublicaEs, Enisa, mitmagob,
 consumogob, Teresaribera, Adif_es, Renfe, redpuntoes, DelGobVG, CienciaGob,
 maec, culturagob, CSIC, deportegob, haciendagob, vsocialgob, SaludISCIII, spain,
 DGTes, salvamentogob, jmrdezuribes, museodelprado, comisionadoPI, EU_Commission,
 museoreinasofia, MuseoThyssen, empleo_SEPE, CDTIoficial, CruzRojaEsp, _CARITAS,
 Cermi_Estatal, AEMPSGOB, UN, OECD, rtve, Congreso, WHO, Sanidadgob, Dragonsaulo,
 CineICAA, LuisPlanas, IgualdadGob, alertcops, CEOE_ES, cepyme_, UGT_Comunica, CCOO,
 SE_Comercio, camarascomercio, Aebanca, sec, TwitterEspana, CREASGR, EUCouncil, IGNSpain,
 CNB_CSIC, osiseguridad, dsn, es_INE, MarotoReyes, Faconauto_com, fempcomunica,
 platdeinfancia, SEDIAgob, ECDC_EU, Vinxco, javidearnedo, josemdomenech2, astefanov5,
 s, Ineco_es, NormasUNE, SaludISCII, Sani, SanidadGob, AndaluciaJu, hans_kluge,
 WHO_Europe, aena, cienciagob, AlertCops, desdelamoncloa, idae, ve, AranchaGlezLaya,
 micoturgob, la2_tve, Clan_tve, UMEgob, 060gobes, SaludMadrid, EjercitoTierra,
 feriademadrid, CrueUniversidad, boe, mscbs, EjercitoTierr, M_Presidencia, defensagob,
 EmbEspanaRabat, astro_duque, GiuseppeConteIT, autonomosata, ilo, EspanaGlobal,
 jensspahn, olivierveran, Fridays4future, HablamosdEuropa, joaquinaraujo,
 AnfacAutomovil, AEMET_Esp, OITnoticias, ThierryBreton, lariojaorg, ConchaAndreu,
 sanidadgo, abalosmeco, Co, mjmonteroc, realessitios, InjuveSpain, 112canarias, ENAIRE,
 GranCanariaCab, CabildoTenerife, PlanTIFIES, AgEInves, Inmujer, Oceanicas_IEO,
 AnimalesGob, Armada_esp, EjercitoAire, congreso_Es, Agenda2030Gob, opengovpart,
 IDAEenergia, usal, iaa_csic, imb_cnm, ONU_es, MWCapital, EmbSpainUK, UNESCO,
 FCBfemeni, MSCActions, CarolinaDarias, CelaaIsabel, carmencalvo_, antoniobanderas,
 Red_Carolina, Turespana_, nuria, diba, nuriamarinlh, FomentTreball, eucopresident,
 CarrefourES, 112cmadrid, sepiegob, EUErasmusPlus, AEPD_es, Twitter, SpainWpFem,
 SpainWP, marcoaguiriano, Agenda2030Esp, EmbEspChina, maecgob, Europarl_ES,
 FBiodiversidad, astro_d, es, RFEBalonmano, san, is4k, JLambanM, TheEconomist, wef_es,
 fitur_madrid, La1_tve, 24h_tve, rne, educaINTEF, CMNUCC, InfoAdif, GVA112, 112Aragon,
 redrunacional, CEAPA3, PNSDgob, AECID_es, proyectoMusaE, educaINEE, IRENA, ATPCup,
 Permafrost_UAH, UNFCCC, fomentogob, el_pais, EU_Careers, Inforenfe, PuertosEstado,
 museothyssen, ESA_CHEOPS, esa, Refugees, FECYT_Ciencia, jcyl, DGCyL, ipcepatrimonio}
```

In[ ]:=
```
SpainTweets[Flatten /* WordCloud, "Usermentions"]
```
aplana nube de palabras

Out[ ]=

Due to the rate limits on the number of calls to the Twitter API we can not download the users data directly. For this reason we have to bypass the API:

In[ ]:=
```
numRequests = 180;
intervals = Ceiling[Length[MentionUsers] / numRequests]
```
entero s··· longitud

Out[ ]= 2

getTwitterData is a function that allow us to donwload UserData from Twitter, given a list of user-Names from the mentionedUser list:

```
getTwitterData[request_String, userNames_List] :=
  twitter[request, "Username" → #] & /@ userNames;
```

Setting up a scheduled task to download user data across multiple rate-limited windows:

In[ ]:=
```
i = 1;
data = {};
SessionSubmit[
```
envía en sesión
```
    ScheduledTask[
```
tarea programada
```
      AppendTo[data,
```
añade al final
```
       getTwitterData["UserData", Partition[MentionUsers, UpTo[numRequests]][[i++]]],
```
particiona hasta
```
       {Quantity[15, "Minutes"], intervals}],
```
cantidad
```
 HandlerFunctions → <|"TaskStatusChanged" → (Print["Task status: ", #TaskStatus] &)|>,
```
funciones de gestión escribe
```
 HandlerFunctionsKeys → {"TaskStatus"}]
```
llaves de funciones de gestión

```
Task status: Running
```

In[ ]:=
```
TaskObject[
```
objeto tarea

Task UUID: 197b2c8e-5131-46d6-bcdc-5a8f7a66e060
Task environment: Session
Task type: Scheduled
Evaluation expression:

```
AppendTo[data, getTwitterData[UserData, Partition[MentionUsers, UpTo[numRequests]][[i++]]]]
```

In[ ]:=
```
MentionedUserData = data;
CleanMentionedUserData =
  Dataset[DeleteCases[Flatten[MentionedUserData], Except[_Association]]];
```
conjunto··· elimina casos aplana excepto

In[ ]:=
```
RandomSample[CleanMentionedUserData]
```
muestra aleatoria

Out[ ]=

ID	ScreenName	Name
1 010 735 383	educaINEE	Educación INEE
20 646 711	UNESCO	UNESCO
21 436 960	esa	ESA
16 941 469	nuria	Nuria
398 574 619	wef_es	WEF en Español
1 068 067 239 700 647 936	marcoaguiriano	Marco Aguiriano
2 350 951 542	JLambanM	Javier Lambán
4 163 160 257	Inforenfe	InfoRenfe
261 225 455	CarolinaDarias	Carolina Darias
125 391 822	usal	Universidad de Salamanca
222 241 634	112cmadrid	112 Comunidad Madrid
57 981 270	La1_tve	La 1
29 977 367	rne	Radio Nacional
803 578 472 929 128 448	InfoAdif	INFOAdif
17 463 923	UNFCCC	UN Climate Change
39 321 874	24h_tve	24h
106 430 178	educaINTEF	Educación INTEF
196 994 616	eucopresident	Charles Michel
889 240 064	jcyl	Junta de Castilla y León
522 128 259	MAECgob	Exteriores

showing 1–20 of 73

We can also visualize the location of each mentioned user with the function GeoListPlot

In[ ]:=
```
locations =
  Interpreter["ComputedLocation"] /@ Normal@CleanMentionedUserData[All, "Location"];
```
intérprete normal todo

In[ ]:=
```
GeoListPlot[DeleteCases[locations, Except[GeoPosition[{_, _}]]]]
representació··· elimina casos    excepto  posición geográfica
```

Out[ ]=

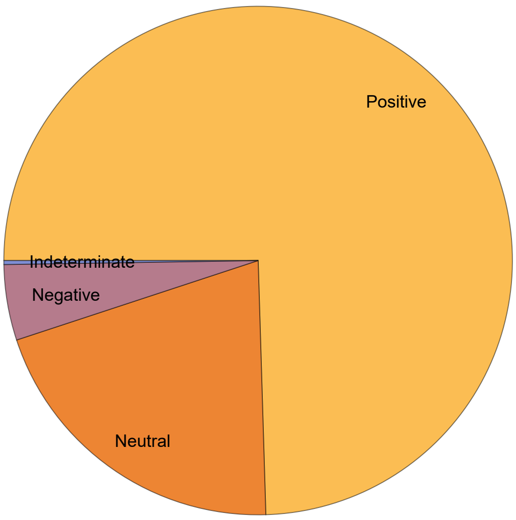
We can use the built-in function “Sentiment” to classify the sentimentes of the tweets posted by @desdelamoncloa. This classifier catalogue each tweet as Positive, Negative, Neutral or Indeterminate.

In[ ]:=
```
labels = Classify["Sentiment", SpainTweets[All, "Text", CleanText]];
         clasifica                          todo  texto
labels[[1 ;; 10]]
```

Out[ ]=
```
{Positive, Positive, Positive, Positive,
 Positive, Positive, Positive, Neutral, Positive, Positive}
```

In[ ]:=
```
PieChart[Counts[labels], ChartLabels → Placed[Automatic, "RadialOuter"]]
diagrama ···conteos      etiquetas de dia··· colocado automático
```

Out[ ]=



As we can see, “Negative” and “Indeterminate” tweets are unlikely, so let us check their contents:

In[ ]:=
```
Cases[({#1 → Classify["Sentiment", CleanText[#1]]} &) /@ SpainTweets[All, "Text"],
  {_ → "Negative"}]
```
casos clasifica todo texto negativo

Out[ ]=
```
"Este #RDL incluye por primera vez una prestación extraordinaria para los a
"✉Las personas trabajadoras de las empresas en #ERTE no podrán realizar h
"❗Las empresas que se acojan a las exoneraciones de los #ERTE, deberán man
"Hoy han reabierto sus puertas los hoteles y restaurantes de @paradores.\n
"Es fundamental impulsar una cultura de Seguridad Nacional entre la ciudada
"ORGULLO \nSigamos reivindicando la diversidad y la visibilidad LGTBI\n\r
"Lograr la igualdad real y efectiva de las personas LGTBI y la no discrimin
" Uso obligatorio de la mascarilla\n識 Prevención e higiene en espacios púl
"Cada día se detectan tres nuevos casos de #ELA en España.\n \nEste Gobiern
"Hay más de 26 millones de personas en el mundo que son refugiadas y al men
"➡Este Plan de Impulso al turismo está dotado con 4.262 millones €\n\nDur
"Hoy se publica en el @boegob la aprobación del Fondo #COVID19.\n\n➡Dotadc
"✅ El @Congreso_Es ha aprobado la tramitación del Proyecto de Ley que derog
"RT @sanchezcastejon: Proteger a la infancia es prioritario, es nuestra obl
"❓Cuándo es obligatorio el uso de la mascarilla en la Nueva Normalidad\n\ı
"Las clases comenzarán en las fechas habituales de septiembre\nLa activic
"✅El Pleno del @Congreso_Es ha convalidado hoy el #IngresoMínimoVital. Una
"Los archivos de @culturagob reanudan hoy la atención presencial en sala. \
"❓Ante la situación económica y social derivada de la crisis sanitaria del
"Esta crisis ha mostrado la fortaleza del sistema de bienestar y las instit
```
showing 1–20 of 155