

PEC 2
Familiarizarte con los datos

Gisele Guadalupe Almeida dos Santos Maia
Marzo 2020.

Universitat Oberta de Catalunya.
Máster de Ciencia de Datos (Data Science).
Visualización de datos

- **Título de la visualización/nes que presentan el dataset escogido.**

Acuerdos de paz Mundial

- **Descripción corta del documento y de los que se presenta.**

El dataset escogido es el de peaceagreements.org, donde tenemos datos de documentos sobre acuerdos de paz en todo el mundo desde el 1990 hasta hoy, descargado en Excel con el nombre del archivo: *pax_all_agreements_data.xlsx*.

El dataset contiene 1.832 líneas y 265 columnas.

Descripción de las columnas:

Con: País de origen del conflicto (territorial)

Contp: Tipo de conflicto: Government, Territory, Government/territory, Inter-group u Other

PP: Código numérico del proceso de paz

PPName: Nombre del proceso de paz

Reg: Región en la que se produce el conflicto abordado por el acuerdo

AgtId: Número de identificación único para cada acuerdo

Agt: Nombre del acuerdo

Dat: Fecha en que se firmó o acordó el acuerdo

Status: Definición y estado del acuerdo: Multiparty signed/agreed, Unilateral agreement, Status unclear, Agreement with Subsequent Status

Lgt: número de páginas del acuerdo

N_characters: número de caracteres en el texto del acuerdo

Agtpt: tipo de acuerdo / conflicto, describe la naturaleza primaria del acuerdo y el conflicto: Inter, InterIntra, Intra, IntraLocal

Stage: Etapa de acuerdo: Pre, SubPar, SubComp, Imp, Ren, Cea y Other

StageSub: subcodificación de etapa

Part: Las partes que firman el acuerdo, las partes en conflicto y sus aliados regionales

ThrdPart: terceros que firman el acuerdo como testigos

OthAgr: Otros nombres de acuerdos de que el documento en cuestión menciona

Loc1ISO: Ubicación primaria del conflicto dentro del estado

Loc2ISO: En casos de conflicto interestatal, indicando una de las ubicaciones principales de conflicto intraestatal

Loc1GWNO: la ubicación primaria del conflicto dentro del estado

Loc2GWNO: en casos de conflicto interestatal, indicando una de las ubicaciones principales

UcdpCon: ID de conflicto, de 2017 en adelante

UcdpAgr: ID de acuerdo de paz

PamAgr: ID de acuerdo de la matriz de acuerdos de paz del Instituto PamAgr Kroc

CowWar: Correlatos de guerra: Número de guerra

En las variables abajo, tenemos grupos, donde 0 no hay referencia o 1 hay referencia. Algunas

variables pueden tener 0, 1, 2 o 3 donde 0 no hay referencia, 1 hay mención, 2 el acuerdo contiene disposiciones de referencia y 3 los derechos están presentes, son substanciales y dan detalles.

Hay también subgrupos de asunto, por ejemplo:

las variables que empiezan con GCh hacen referencia a los Niños y Jóvenes y cada una de las demás a uno asunto en relación con ellos y estos asuntos se repiten.

Grupo: GCh: Niños / Jóvenes

GChRhet: compromisos preambulares u otros retóricos

GChAntid: compromiso o una disposición contra la discriminación

GChSubs: mecanismos integrales o compromisos

GChOth: el acuerdo proporciona otras disposiciones

Así, haremos solamente la descripción de los

Grupos ya que las variables XXxRhet, XXxAntid, XXxSubs y XXxOth son las mismas relacionadas con el grupo que pertenecen.

GDis: personas discapacitadas

GAge: personas mayores

GMig: trabajadores migrantes

GRa: acuerdo de raza, grupos étnicos, minorías nacionales, clanes u organizaciones similares 'tribales'

GRe: grupos religiosos

GInd: pueblos indígenas

GOth: otros grupos

GRef: refugiados, personas desplazadas

GSoc: clase social u origen social

Las variables que empiezan con Ge son relacionadas a Género

GeWom, GeMe, GeLgbti. GeFa

Las variables que empiezan con St son relacionadas con Definición del Estado:

StDef; StGen; StCon; StSd; StRef; StSym; StInd; StUni; StBor; StXbor

Las variables relacionadas a Gobernancia: PolGen; PolNewInd; PolNewTemp; ConRen; Cons; Ele; ElecComm; PolPar; PolParTrans; PolParOth; Civso; Tral; Pubad

Variables relacionadas a Reparto del poder: Polps; PpsSt; PpsSub; PpsEx; PpsOro; PpsOthPr; PpsVet; PpsAut; PpsInt; PpsOth; Terps; TpsSub; TpsLoc; TpsAut; TpsOth; Eps; EpsRes; EpsFis; EpsOth; Mps; MpsMe; MpsJt; MpsPro; MpsOth

Variables relacionadas con Derechos humanos e igualdad: HrGen; EqGen; HrDem; Prot; ProtCiv; ProtGrp; ProtLgl; ProtOth; HrFra; HrfSp; HrfBor; HrfTinc; HrfOth; HrCp; CprLife; CprTort; CprSlav; CprLib; CprDet; CprFmov; CprFspe; CprFass;

CprTria; CprPriv; CprVote; CprReli; CprOth; HrSec; SerProp; SerWork; SerHeal; SerEdu; SerStdI; SerShel; SerSs; SerCult; SerOth; HrNi; HrNiMe; HrNiNe; HrNiOth; Hrli; HrliMon; HrliBod; HrliOth; HrMob; HrDet; Med; MedGov; MedSubs; MedLog; MedOth; HrCit; CitGen; CitRights; CitDef; CitOth

Variables relacionadas con Reforma del sector de justicia: JusCr; JusCrSp; JusCrSys; JusCrPow; JusEm; JusJu; JusPri; JusTra;

Variables relacionadas con Reconstrucción socioeconómica: Dev; DevSoc; DevHum; DevInfra; NEC; NatRes; IntFu; Bus; Tax; TaxPo; TaxRef; TaxOth; Ban; CenBan; BanPers; BanInt; BanXb;

Variables relacionadas con Tierra, propiedad y medio ambiente: LaRef; LaRefMan; LaRefRet; LaRefOth; LaNom; LaCH; LaCHTa; LaCHIt; LaCHPro; LaCHOth; LaEn; Wat

Variables relacionadas con Sector de seguridad: SsrGua; Ce; CeProv; CeGen; SsrPol; SsrArm; SsrDdr; DdrDemil; DdrProg; SsrInt; SsrPsf; SsrFf; Cor; SsrCrOcr; SsrDrugs; Terr

Variables relacionadas con Justicia transicional: TjGen; TjAm; TjAmPro; TjSan; TjPower; TjAmBan; TjCou; TjJanc; TjJalc; TjMech; TjPrir; TjVet; TjVic; TjMis; TjRep; TjRSym; TjRMA; TjNR; ImUN; ImOth; ImRef; ImPK; ImE; ImSrc;

- **Qué he aprendido explorando los datos: tipo, estructura, curiosidades**

Mirando los duplicados por Código numérico del proceso de paz (PP) tenemos 156 procesos distintos de paz.

```
## Mirando los distintos procesos de paz (PP)
base.PP.nunique()
```

156

Hay solamente 12 variables con registros nulos: StageSub, Part, ThrdPart, OthAgr, Loc1ISO, Loc2ISO, Loc1GWNO, Loc2GWNO, UcdpCon, UcdpAgr, PamAgr, CowWar.

La con más registros nulos es la variable PamAgr con 98,2% de missing y la con menos es: Loc1GWNO con 0,08%.

```
## % de missing
percent_missing = base.isnull().sum() * 100 / len(base)
percent_missing
```

Lgt	0.000000
N_characters	0.000000
Agtp	0.000000
Stage	0.000000
StageSub	1.255459
Part	0.873362
ThrdPart	48.307860
OthAgr	47.598253
Loc1ISO	2.074236
Loc2ISO	87.227074
Loc1GWNO	0.764192
Loc2GWNO	87.227074
UcdpCon	7.696507
UcdpAgr	83.788210
PamAgr	98.198690
CowWar	39.246725
GCh	0.000000
GChRhet	0.000000

Tenemos solamente 16 variables de texto, las otras 249 son numéricas. La variable Dat que es la fecha en que se firmó o acordó el acuerdo tiene el formato de string, así, es necesario cambiarla al formato de fecha.

- **Qué habéis descubierto: evidencias, tendencias, outlayers, etc.**

La mayoría de las variables tienen valores de 0 o 1, de 0 hasta 3 según vimos en la descripción de las variables.

```
base.describe().transpose()
```

	count	mean	std	min	25%	50%	75%	max
PP	1832.0	8.395360e+01	4.180348e+01	1.0	53.00	90.0	120.00	1.580000e+02
Agtp	1832.0	1.132910e+03	6.508488e+02	1.0	557.75	1110.5	1704.75	2.285000e+03
Lgt	1832.0	6.289301e+00	1.730743e+01	1.0	1.00	2.0	5.00	3.230000e+02
N_characters	1832.0	1.184655e+04	3.716447e+04	329.0	2382.00	4345.0	9377.50	9.084590e+05
Loc1GWNO	1818.0	5.288091e+02	2.252744e+02	41.0	365.00	531.0	678.00	9.400000e+02
Loc2GWNO	234.0	4.372778e+02	1.984422e+02	2.0	345.00	371.0	625.00	7.700000e+02
UcdpAgr	297.0	2.903701e+10	3.602522e+11	24.0	1107.00	1334.0	1523.00	5.171991e+12
PamAgr	33.0	1.727273e+01	1.002270e+01	1.0	9.00	17.0	26.00	3.400000e+01
CowWar	1113.0	7.602848e+02	2.539813e+02	139.0	817.00	857.0	905.00	9.410000e+02
GCh	1832.0	3.155022e-01	7.714314e-01	0.0	0.00	0.0	0.00	3.000000e+00
GChRhet	1832.0	8.187773e-02	2.742532e-01	0.0	0.00	0.0	0.00	1.000000e+00

GChAntid	1832.0	3.275109e-03	5.715038e-02	0.0	0.00	0.0	0.00	1.000000e+00
GChSubs	1832.0	1.048035e-01	3.063837e-01	0.0	0.00	0.0	0.00	1.000000e+00
GChOth	1832.0	1.091703e-02	1.039411e-01	0.0	0.00	0.0	0.00	1.000000e+00
GDis	1832.0	1.375546e-01	5.546764e-01	0.0	0.00	0.0	0.00	3.000000e+00
GDisRhet	1832.0	2.074236e-02	1.425595e-01	0.0	0.00	0.0	0.00	1.000000e+00
GDisAntid	1832.0	9.825328e-03	9.866156e-02	0.0	0.00	0.0	0.00	1.000000e+00
GDisSubs	1832.0	4.803493e-02	2.138985e-01	0.0	0.00	0.0	0.00	1.000000e+00
GDisOth	1832.0	2.729258e-03	5.218520e-02	0.0	0.00	0.0	0.00	1.000000e+00
GAge	1832.0	9.115721e-02	4.469551e-01	0.0	0.00	0.0	0.00	3.000000e+00
GAgeRhet	1832.0	1.801310e-02	1.330349e-01	0.0	0.00	0.0	0.00	1.000000e+00
GAgeAntid	1832.0	7.641921e-03	8.710720e-02	0.0	0.00	0.0	0.00	1.000000e+00
GAgeSubs	1832.0	2.947598e-02	1.691827e-01	0.0	0.00	0.0	0.00	1.000000e+00
GAgeOth	1832.0	0.000000e+00	0.000000e+00	0.0	0.00	0.0	0.00	0.000000e+00

Los tipos de conflictos que más tenemos acuerdos es: Government/territory con 45,6% de los registros y luego Government con 35,6%

```
contp = pd.crosstab(index=base['Contp'],
                    columns='%')
(contp/contp.sum())*100
```

col_0	%
Contp	
Government	35.644105
Government/territory	45.578603
Inter-group	12.827511
Other	0.054585
Territory	5.895197

La región con más acuerdos de paz es Africa (excl. MENA) con 30,2% de los acuerdos

```
reg = pd.crosstab(index=base['Reg'],
                  columns='%')
(reg/reg.sum())*100
```

col_0	%
Reg	
Africa (excl MENA)	30.240175
Americas	10.644105
Asia and Pacific	20.578603
Cross-regional	1.855895
Europe and Eurasia	22.379913
Middle East and North Africa	14.301310

Los años con más acuerdos de paz fueran en 1993 y 1994

```
year = pd.crosstab(index=base['year'],
                    columns='%')
(year/year.sum())*100
```

col_0	%
year	
1990	2.401747
1991	4.585153
1992	4.530568
1993	5.021834
1994	5.076419
1995	3.548035
1996	3.602620
1997	3.930131
1998	3.602620
1999	4.639738
2000	3.275109
2001	3.165939
2002	2.893013
2003	2.783843
2004	2.565502
2005	1.637555
2006	3.329694
2007	2.401747
2008	3.493450
2009	2.510917
2010	1.801310
2011	2.783843
2012	3.384279
2013	2.292576
2014	4.312227
2015	3.766376
2016	4.093886
2017	3.820961
2018	3.711790
2019	1.037118

92% de los acuerdos tienen la definición y estado del acuerdo cómo Multiparty signed/agreed

```
Status = pd.crosstab(index=base['Status'],
                      columns='%')
(Status/Status.sum())*100
```

	col_0	%
Status		
Agreement with subsequent status	0.382096	
Multiparty signed/agreed	92.303493	
Status unclear	0.545852	
Unilateral document	6.768559	

69,5% de los acuerdos son del tipo: Intra - conflictos dentro de las fronteras de un estado

```
Agtp = pd.crosstab(index=base['Agtp'],
                   columns='%')
(Agtp/Agtp.sum())*100
```

	col_0	%
Agtp		
Inter	3.930131	
InterIntra	12.882096	
Intra	69.486900	
IntraLocal	13.700873	

Las etapas que tenemos más acuerdos son: Pre – Prenegociación (27,8%) , SubPar – marco sustantivo, parcial (24.9%) y Cea - alto el fuego / relacionado (20,8%)

```
Stage = pd.crosstab(index=base['Stage'],
                    columns='%')
(Stage/Stage.sum())*100
```

	col_0	%
Stage		
Cea	20.796943	
Imp	18.067686	
Oth	0.054585	
Pre	27.838428	
Ren	2.292576	
SubComp	6.004367	
SubPar	24.945415	

- **Explicar cómo lo habéis descubierto: con qué herramientas, cálculos , con qué operaciones**

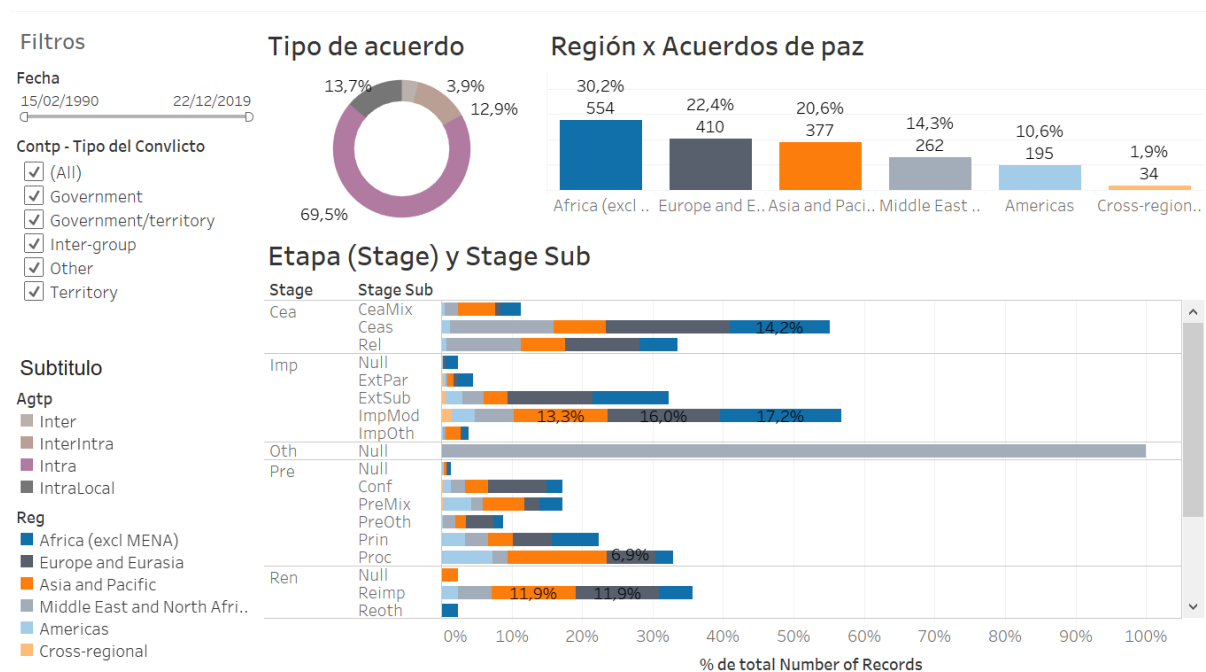
Para la exploración de los datos he utilizado Python, la librería Pandas para mirar los formatos, cambiar lo formato de la fecha, hacer la distribución porcentual de algunas variables en relación al número total de los cuerdos y, también para mirar cantidad, máximo, mínimo, promedio, algunos cuartiles y desviación estándar con la función: “describe()”.

En el código de las figuras de los cálculos, tenemos que “base” es la base de datos descargado en la página web: peaceagreements.org, de nombre: *pax_all_agreements_data.xlsx*.

- **Visión general de las diversas visualizaciones coordinadas en un dashboard**

URL:

<https://public.tableau.com/profile/gose.e.gudalupe.almeida.dos.santos#!/vizhome/AcuerdosdePazMundial/AcuerdosdePazMundial>



Filtros

Fecha

15/02/1990

22/12/2019

Contp - Tipo del conflicto

- ☒ (All)
- ☒ Government
- ☒ Government/territory
- ☒ Inter-group
- ☒ Other
- ☒ Territory

Subtítulo

Number of Records

7 1.691

Status

Multiparty signed/agreed
92,3%
1.691

Mapa de Acuerdos de Paz

