

# PAC 2. Visualización de Datos.

## 1. Título.

Comparativa temporal, geográfica y temática de los 10 procesos de paz con más acuerdos.

## 2. Descripción del documento y de lo que se presenta.

En este documento de la PAC 2, se ha estudiado la base de datos de acuerdos de paz de <https://www.peaceagreements.org/search>. En primera instancia se ha estudiado la composición de los datos de manera cualitativa, identificando las variables clave y comprendiendo la distribución del dataset entre información general y aspectos de los acuerdos.

A continuación, se ha estudiado el tipo de las variables, identificando tanto aquellas categóricas como numéricas, mencionando los errores de identificación de Python y mencionando las cualidades que pueden tener las variables. Por último, para la tarea 3 se han realizado unas pequeñas transformaciones que nos permiten posteriormente explorar el número de acuerdos por país y la temática de los acuerdos.

En la tarea 4 se ha explorado el dataset mediante Python, se han descubierto evidencias claras en los datos como que el país con más acuerdos es Yugoslavia o que la mayoría de acuerdos son de tipo intraregional. A continuación, se realiza una exploración más profunda y mediante el estudio de las variables de tipo de conflicto, acuerdo y temática de grupos para cada proceso de paz del top 10, se ha descubierto que los procesos son diferentes entre ellos y que los acuerdos de paz internos a un proceso son similares entre ellos. También se ha estudiado aspectos curiosos como que los acuerdos y procesos que tratan un tema tienen correlación a tratar otros temas, que hay acuerdos dedicados a tratar temas muy concretos que se encuentran como outliers y que la tendencia de los acuerdos de paz por año es negativa.

En la tarea 5 se ha descrito el procedimiento realizado para obtener estas conclusiones de la tarea 4, se ha explicado el software utilizado para la exploración y se han expuesto las diferentes posibilidades de uso para desarrollar la visualización de la tarea 6, marcando una decisión final por Tableau.

En la tarea 6 se ha realizado el dashboard mencionado con el objetivo de caracterizar y comparar el top 10 de los procesos de paz de manera temporal, geográfica y temática para encontrar diferencias y similitudes entre ellos. El dashboard presenta una visión general, seguida de una específica para esta comparativa y se aprovecha del color usado y de la interactividad para facilitar la comprensión y uso de quién lo visualice. En la tarea 6 se menciona todo lo realizado en el dashboard.

El autor ha sido Manuel Ruiz Botella, alumno del Máster Universitario de Ciencia de Datos por la UOC. Ha sido terminado a fecha de 20 de Abril de 2021, en Tarragona. La razón principal es para la realización de la PAC 2 de la asignatura de Visualización de Datos del Máster, aunque con el desarrollo de la PAC se ha convertido en razón principal el interés por descubrir las diferencias y parecidos entre los procesos de paz más importantes desde los 90. Como ya se ha mencionado, la base de datos se puede encontrar en: <https://www.peaceagreements.org/search>.

Los archivos generados durante el desarrollo se adjuntan en la entrega y se encuentran subidos en: [https://github.com/manurubo/PAC2\\_VD\\_Manurubo](https://github.com/manurubo/PAC2_VD_Manurubo)

### 3. Los datos, presentación. ¿Qué sabéis de los datos? tipo, estructura, curiosidades.

#### **Presentación cualitativa:**

En esta PEC, vamos a utilizar la base de datos de acuerdos de paz. Esta base de datos contiene 1868 acuerdos de paz encontrados en más de 150 procesos de paz entre 1990 y 2020. La base de datos tiene varias versiones, siendo la 4ª la que vamos a utilizar en esta PEC y la más reciente. La base de datos se encuentra en: <https://www.peaceagreements.org/search>

La base de datos contiene diversos tipos de información, por una parte, contiene información general o básica de los acuerdos. Las variables más importantes de manera general que se encuentran hacen referencia a los países que participan en el conflicto, al tipo de conflicto (territorial/gubernamental e inter/intra), al nombre del proceso de paz que pertenece el conflicto, a la región donde se realiza un acuerdo, al status del acuerdo (unilateral o multiparte), a la etapa del proceso de paz a la que pertenece un acuerdo o a las partes de un acuerdo.

El resto de los datos de la base pertenecen a características del acuerdo de paz, la base de datos contiene información sobre si un acuerdo contiene palabras respecto a:

- **Los grupos de personas.** Ya sean niños, discapacitados, ancianos, inmigrantes, razas, religiones, indígenas, refugiados, clases sociales o otros. Para cada grupo hay subcategorías en función de la importancia que le da un acuerdo al tema.
- **Los géneros de las personas.** Determinando si se mencionan mujeres, hombres, términos LGTBI o de familia.
- **La definición de un estado.** Y las submaneras en las que se puede definir un estado (autodeterminación, referéndum, independencia, unificación) y aspectos importantes de la definición de un estado, como su naturaleza, geografía, símbolos, fronteras y comunicación con otros países.
- **El gobierno de un estado.** Interpretando si en el acuerdo se tratan las instituciones políticas, una constitución nueva o una reforma, la realización de elecciones o una comisión electoral, la reorganización de partidos políticos, una sociedad civil, los líderes tradicionales, o la administración pública.
- **La compartición de poderes.** Ya sea una compartición de los poderes políticos, territoriales, económicos o militares. Dentro de cada compartición de poderes, se tratan los aspectos claves de cada uno en diferentes variables, por ejemplo respecto a la compartición de poderes territoriales, se contienen subvariables por si hay compartición federal, local, de autonomía o otra.
- **Los derechos humanos y la igualdad.** Donde se busca que un acuerdo de paz hable sobre los derechos humanos, la igualdad, la democracia, medidas de protección para diversas categorías entre ellas los civiles o grupos en desigualdad, un marco de declaración de derechos más o menos formado, derechos civiles y políticos en diferentes ámbitos como la vida o la esclavitud entre otros, derechos socioeconómicos como la pobreza o la salud entre otros, si menciona instituciones de los derechos humanos nacionales o internacionales, si trata la movilidad de las personas, los procesos

de detención, o la libertad que tienen medios de comunicación, si trata de establecer una ciudadanía y lo que representa ser ciudadano respecto a derechos.

- **Una reforma en el sector jurídico.** Centrándose en el sistema criminal de justicia, las provisiones en un estado de emergencia, las cortes y el poder judicial, la prisión y las detenciones o las leyes tradicionales o clásicas.
- **Una reconstrucción socioeconómica.** Ya sea tratando el desarrollo o reconstrucción socioeconómica, un plan económico nacional, los recursos naturales, las ayudas internacionales, las ayudas a negocios, los impuestos o los bancos.
- **La tierra, propiedad y el entorno.** Las variables tratan si se habla sobre reformas en la tierra o derechos sobre esta, si se incluyen derechos para tribus/pastores, si hay herencia cultural tangible intangible o su promoción, si se trata del entorno ecológico o si se los acuerdos mencionan el acceso al agua.
- **El sector de seguridad.** Se tratan los diferentes aspectos respecto a la seguridad como la garantía de seguridad, los altoelfuego, la policía, las fuerzas armadas, la vuelta a una vida normal/desmilitarizada, los servicios de inteligencia, los grupos de la oposición, el retiro de fuerzas externas, la corrupción, el crimen organizado, las drogas, o el terrorismo.
- **La justicia transicional.** Donde se marcan los aspectos claves de la transición a la paz, como la complejidad de la justicia general para la transición, la amnistía, la formación de cortes, el mecanismo para la transición, la liberación de prisioneros, los vetos, las asistencias a las víctimas, la información sobre muertos y perdidos, las provisiones de reparaciones y las posibles reconciliaciones.
- **Implementación.** Se notifica si las naciones unidas firman el acuerdo, u otro firmante internacional, si el acuerdo se va a llevar a referéndum, si son misiones internacionales y mecanismos de refuerzo del acuerdo.

#### **Presentación cuantitativa:**

Todos estos datos conforman 266 columnas que son atributos de acuerdos de paz para 1868 acuerdos de paz, que son las filas del dataset.

La información general de un acuerdo de paz contiene 26 variables, y el resto de las variables se encuentran distribuidas en grupos como ya hemos visto. Cada grupo tiene un número de variables y subvariables:

- Los grupos de personas contienen 10 grupos, donde cada grupo tiene 4 subvariables salvo el de refugiados que solo tiene 3. Por tanto, hay 49 variables.
- Para los géneros de las personas hay 4 variables, pero tanto la que hace referencia a hombres y niños como la que hace referencia a LGTBI se divide en 2 subvariables, que hacen 8 variables.
- Respecto a la definición de estado, tenemos una variable que indica si se define un estado, y 9 subvariables que indican aspectos de este estado, por tanto en total son 10 variables.
- Al hablar del gobierno de un estado, vemos que hay 9 variables ya comentadas, pero las instituciones políticas que se generan se dividen en identificar si hay referencias generales, si son nuevas instituciones y si son temporales, y la reforma de partidos políticos se divide en si los rebeldes transicionan a partidos políticos y si hay regulaciones de los partidos o reformas, por lo que en total hacen 14 variables.

- Para la compartición de poderes, cada variable de las 4 se divide en varias subvariables, haciendo en total 24 variables.
- Con los derechos humanos e igualdad ocurre lo mismo, algunas variables se subdividen y otras apenas nada, por tanto, tenemos 58 variables diferentes.
- Para la reforma del sector jurídico, nos encontramos que solo la variable de justicia criminal y leyes de emergencia se subdivide en reformas a leyes específicas, reforma al sistema criminal o delimitación de poderes del sistema de justicia, en total hacen 8 variables.
- Al estudiar la reconstrucción socioeconómica, nos encontramos que hay varias variables como la reconstrucción y desarrollo, los impuestos o los bancos que se subdividen. Hacen en total 17 variables.
- Al hablar de tierras, propiedad y entorno ocurre igual, haciendo en total 12 variables entre las variables y subvariables.
- El sector de seguridad contiene la mayoría variables sin dividir excepto para los altosel fuego y para la desmilitarización, que se dividen en dos variables en función de la seriedad con la que hay un alto el fuego o una desmilitarización. Hay 16 variables.
- Cuando se estudia la justicia de la transición, tenemos 3 variables que se subdividen, tanto las amnistías, como las cortes y las reparaciones. Haciendo en total 18 variables.
- La implementación son variables simples de aspectos finales de un acuerdo, que se recogen en 6 variables.

#### **Tipos y transformaciones:**

Es importante destacar que los tipos de las variables se han encontrado en el archivo html generado a partir de un fichero jupyter notebook en Python. Aquí se comenta por encima lo que se descubre en el fichero. El archivo con nombre “**Exploración\_Datos\_PEC2\_VD.ypinb**” se encuentra en los ficheros de entrega y en el repositorio git indicado. Este archivo también se utilizará para la exploración de los datos en el apartado 4.

La distribución de variables según Python pandas es la siguiente:

```
int64      245
object     16
float64      5
```

**Vemos que hay muy pocas variables categóricas en el conjunto de datos, solo 16.** Estas son los países del acuerdo, el tipo de conflicto respecto a si es gubernamental o territorial, el nombre del proceso de paz, la región, el tipo del acuerdo y su fecha, el status del acuerdo, el tipo de acuerdo si es intraestado o interestados, la fase del proceso de paz y la subfase, las partes y terceras partes, los otros acuerdos a los que se hace referencia, la codificación ISO de los países y la codificación UCDP.

De estas variables, vemos que **algunas son realmente texto libre** como los participantes o los acuerdos a los que se hace referencia en el acuerdo que se estudia, incluso se podría considerar el nombre del propio acuerdo.

Por otra parte, tenemos la fecha del acuerdo, **en la variable Dat, que realmente es tipo fecha o Date**, aunque en Pandas se interpreta como categórica. No es importante cambiarla, puesto que para la tarea 6, se interpretarán los datos correctamente al usar Tableau.

**Por último, tenemos las variables categóricas realmente del acuerdo**, que son atributos directamente del acuerdo, nos encontramos con las variables respecto a los países del acuerdo,

la región, el proceso de paz y los atributos del acuerdo como la naturaleza del conflicto, del acuerdo, el status o la fase del proceso de paz en el que se encuentra.

Es importante destacar que la mayoría de las variables categóricas ya tienen valores válidos para una exploración de los datos, sin embargo, si estudiamos los países de cada acuerdo en la columna Con, vemos que cada acuerdo (fila) contiene los países que participan en el acuerdo separados por /.

Por ejemplo, si vemos los países que salen más en la columna Con (imagen abajo), nos encontramos con que la primera opción es “Bosnia and Herzegovina/Yugoslavia (former)” y la tercera “Philippines/Mindanao”. Esto hace que, si ponemos el número de acuerdos por país en un mapa, se utilizaría siempre el primer país de la lista y no se contaría el acuerdo para todos los países de la lista de Con en el acuerdo. Por tanto, en el tercero los 126 acuerdos de paz entre Philippines y Mandanao solo valdrían para Philippines.

```
Bosnia and Herzegovina/Yugoslavia (former) 133
Colombia 128
Philippines/Mindanao 126
Syria 93
Nepal 51
South Sudan 51
Yemen 50
South Sudan/Sudan 44
Croatia/Yugoslavia (former) 44
Central African Republic 43
Name: Con, dtype: int64
```

**Aquí sería necesaria una transformación de los datos en caso de que queramos representar realmente los países de los acuerdos de paz. Por tanto, la transformación de datos consistirá en generar una fila para cada acuerdo de paz por país,** para el ejemplo de Philippines y Mandanao se duplicarán los 126 acuerdos, pero cada parte del duplicado solo tendrá un país.

**Para mantener la coherencia, en el fichero de Python se ve como se genera otro archivo csv con el contenido de los países y la variable que identifica un acuerdo de paz. De esta manera, el dataset original conserva su estructura y el nuevo dataset con cada país asociado correctamente en una fila a un acuerdo se podrá unir por el identificador con la información de los acuerdos. Este nuevo fichero se llama “Countries\_and\_agtid.csv”.**

**A continuación, se pueden estudiar las variables float,** que son 5 variables con tipo numérico real. **Las variables Loc1GWNO y Loc2GWNO son variables numéricas** que hacen referencia a un código de localización, y por tanto son asignadas correctamente como **valores numéricos reales**. **Sin embargo, las variables UcpdCon, PamAgr y CowWar son realmente identificadores** que no importan en nuestro estudio **y son categóricas**, pero en Python se interpretan como numéricas reales por los valores nulos.

**El resto de variables son numéricas,** por una parte tenemos el PP y el AgtId, el primero hace referencia a un identificador para los procesos de paz y el segundo para los acuerdos de paz, estas variables realmente pueden considerarse como numéricas o como categóricas, ya que son identificadores. Además, tenemos la versión de la base en la que se añadió el acuerdo, la longitud en páginas y la longitud en caracteres, estas variables son realmente numéricas.

**Por último, tenemos el resto de variables numéricas que determinan características de un acuerdo** (grupos de personas, factores socioeconómicos, judiciales, etc), por una parte **actualmente son variables numéricas porque vienen codificadas con números enteros, pero las variables se pueden interpretar como categóricas puesto que hay variables binarias** donde los valores representan un sí o un no mediante 1/0 y también, **hay variables categóricas**

**ordenadas** que representan la importancia del aspecto que se estudia en la variable del acuerdo mediante un número entero que cuanto mayor sea, mejor se representa esa característica en el acuerdo.

Respecto a estas variables, puede ser interesante realizar transformaciones de agrupación de variables. Por ejemplo, agrupar cada uno de los grupos de variables discretos, en una puntuación máxima final, de esa manera para cada acuerdo tenemos la puntuación de grupos de personas, de grupos de género, de justicia transicional, de seguridad, etc. **Esta agrupación de las variables de cada grupo en una única puntuación nos permitiría obtener unos pocos atributos con los que comparar entre acuerdos o procesos de paz y hacer un perfilado general antes de entrar en detalle en las variables de cada grupo.**

En documento de exploración se encuentra esta transformación, de tal manera que para cada acuerdo de paz se obtiene una puntuación de suma de todas las variables de un tema, estas variables se conocerán como “Total\_§tema” siendo §tema el tema que se trata en la variable. Por ejemplo, la variable “Total\_Groups” es la suma de todas las variables del tema Groups, así, esta variable refleja “la importancia” o cuánto se trata un tema en el acuerdo.

## 4. Los datos, exploración. ¿Qué habéis descubierto? evidencias, tendencias, outliers, etc.

Previo a la realización de una visualización, es necesario explorar los datos para ver que puede ser interesante para estudiar en la visualización de los datos. La exploración de los datos se encuentra en el fichero “Exploración\_Datos\_PEC2\_VD.ypinb”.

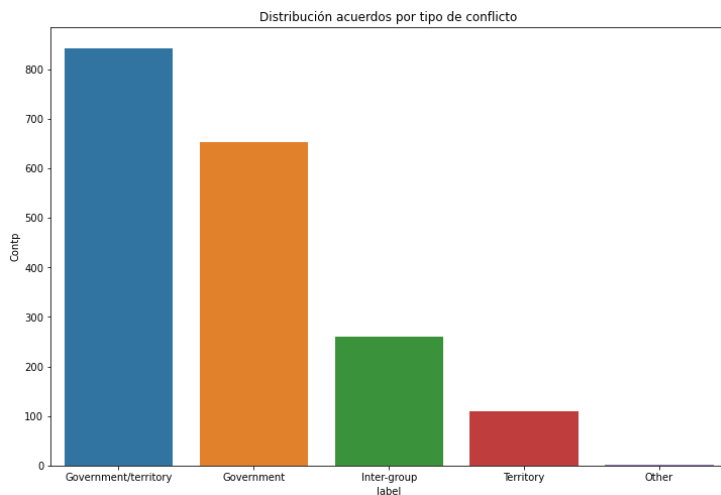
### Evidencias encontradas en los datos.

Durante la exploración, encontramos diferentes aspectos que son evidentes en los datos y que no se tiene que poner mucho esfuerzo en encontrarlos.

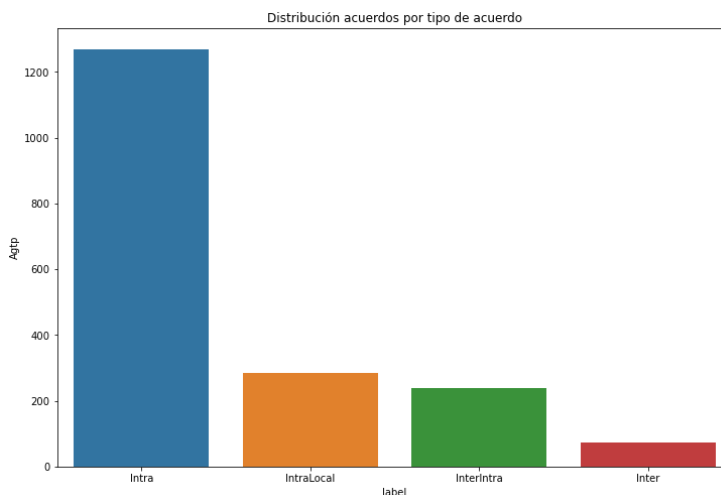
**Una evidencia que encontramos en los datos es el desbalanceo en la distribución de los acuerdos de paz entre países, viendo que hay países/regiones que concentran la mayoría de los acuerdos de paz, y que son zonas donde hay más guerra en este tiempo.** Los países que más concentran estos acuerdos de Paz son la antigua Yugoslavia, Filipinas, Bosnia y Herzegovina, Colombia, Mindanao, Sudan o Siria, todos estos países que claramente son conocidos por todos que durante estos años han mantenido conflictos bélicos. Esto se puede ver reflejado en el Notebook de exploración.

**Otra evidencia encontrada es que los procesos de paz con más acuerdos de datos son lógicos respecto a los países/regiones con más acuerdos de paz, siendo los procesos de paz más importantes el de Filipinas-Mindanao y el de Bosnia (que interviene Yugoslavia).** Además, se encuentran otros procesos de paz que involucran a países ya mencionados. También podemos ver en la exploración de los datos la figura respecto a esto.

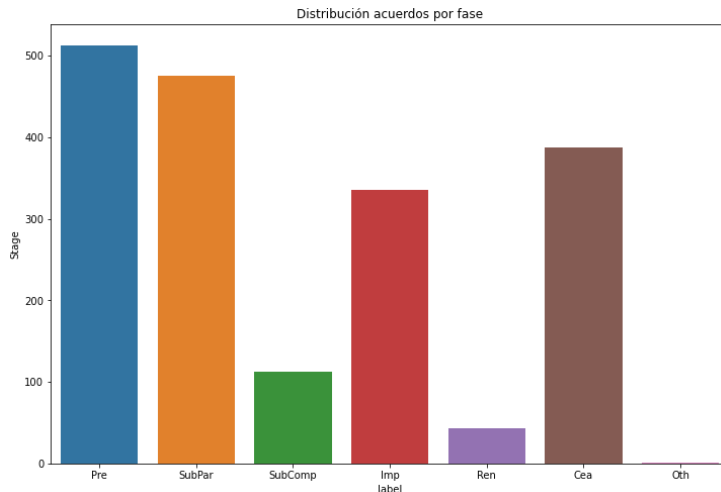
**También encontramos en la exploración que la mayoría de conflictos que ocurre en un acuerdo de paz son de tipo gubernamental, destacando sobre todo aquellos gubernamentales/territoriales.** Hay muy pocos conflictos centrados principalmente en el territorio, habiendo apenas 100 entre los 1868 acuerdos de paz.



También vemos evidente en los datos que la inmensa mayoría de acuerdos, más de 1200 de los 1868 acuerdos, son de tipo intranacionales, seguidos de acuerdos locales y los intrarregionales con agentes internacionales. Los acuerdos que menos se dan son los internacionales, aunque normalmente esos conflictos son los más sonados.



Como última evidencia de los datos, estudiamos los acuerdos por la fase del proceso de paz al que pertenecen. En estos podemos sacar evidencias diferentes. Por una parte, que **la mayoría de los acuerdos ocurren en la fase previa**, donde hay varias partes de un acuerdo intentando la paz, es posible que esta sea la categoría con más acuerdos porque se rompen las negociaciones en las fases posteriores de un proceso de paz. Por otra parte, vemos que **normalmente hay más acuerdos tipo SubPar que SubComp**, esto significa que **normalmente hay más acuerdos de paz donde las partes del acuerdo encuentran discrepancias en algunos aspectos a implementar del acuerdo que acuerdos de paz donde las partes del acuerdo aceptan totalmente el acuerdo entero y sus medidas**.



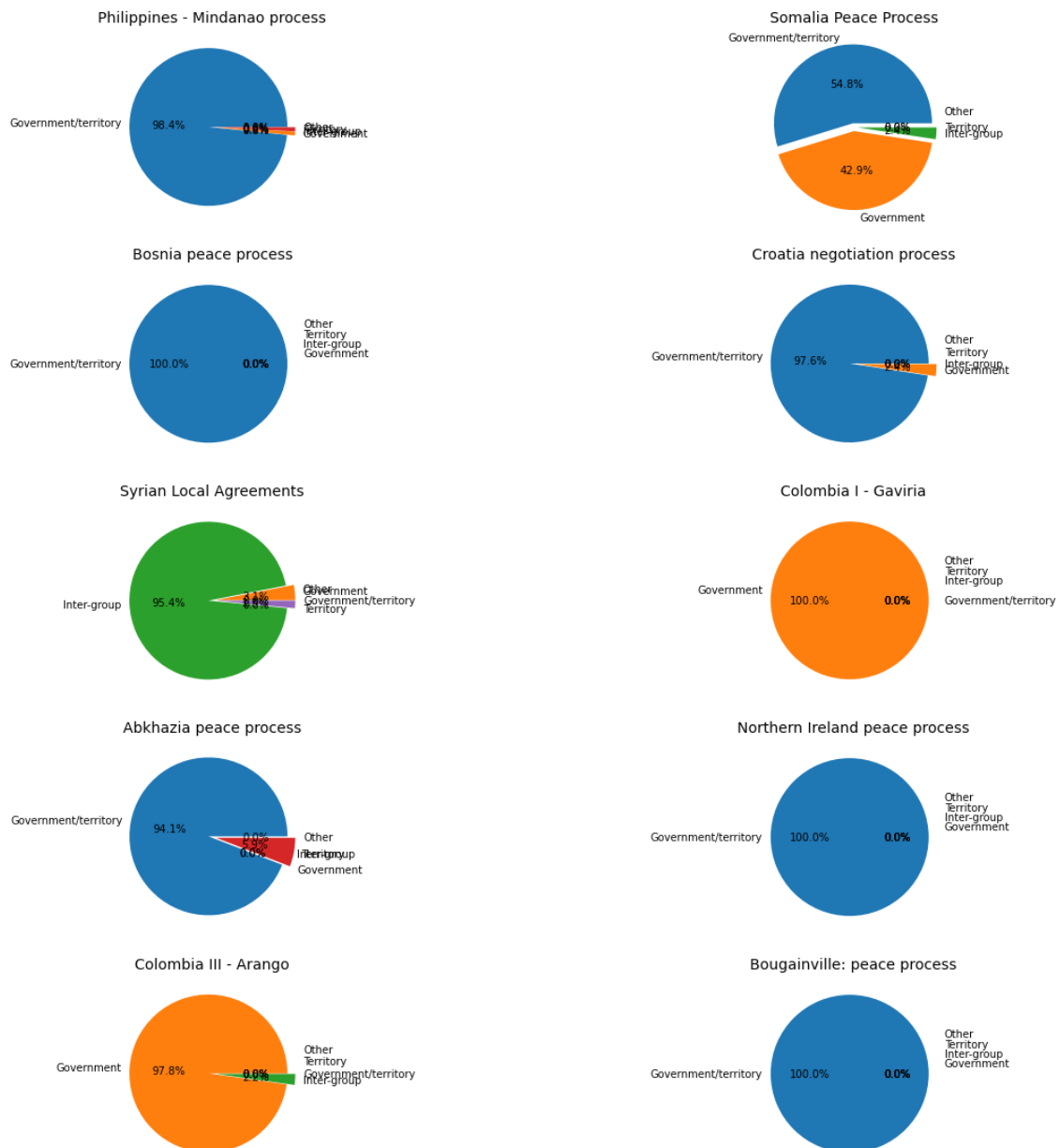
### ¿Qué encuentro potencialmente interesante para realizar una visualización?

Durante la exploración realizada hasta ahora, hemos visto que hay acuerdos de paz y procesos de paz. Los procesos de paz engloban a varios acuerdos de paz que recogen un mismo conflicto o aspectos entre países o regiones. Hemos visto también que los países con más acuerdos de paz suelen ser porque han participado en un proceso de paz con muchos acuerdos.

Puede ser interesante estudiar los procesos de paz con más acuerdos de paz, precisamente porque es lógico pensar que dentro de un proceso de paz habrá acuerdos de paz que traten un mismo conflicto y por tanto habrá similitudes entre sus acuerdos de paz. Esto también nos lleva a pensar que habrá diferencias entre los procesos de paz y deberíamos de poder distinguirlos por sus características.

**En la exploración hemos encontrado que a pesar de que el conjunto entero de acuerdos de paz tiene una distribución de tipo de conflicto ya vista de manera general anteriormente, cuando miramos el porcentaje de tipos de conflicto interno de cada proceso de paz del top 10, nos encontramos con que la distribución no se mantiene.** Al contrario, vemos que normalmente hay un tipo de conflicto que se mantiene mayoritario en un proceso de paz. Normalmente, entre los procesos de paz del top 10, el tipo de conflicto que destaca en los acuerdos es el de Gobierno Territorio. Sin embargo, podemos ver como en el proceso de paz Sirio, la mayoría de acuerdos son Inter-Group y para los acuerdos de Colombia I y Colombia III la mayoría son tipos de conflicto solo de Gobierno.

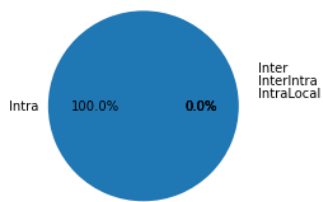




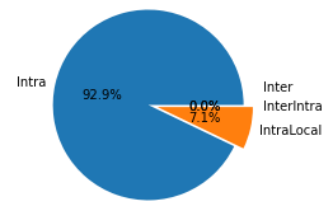
Por tanto, encontramos que hay procesos que se diferencian entre ellos y otros que se parecen entre ellos de acuerdo a esta variable. Lo lógico es que sea por que se tratan conflictos diferentes en cada proceso, y los acuerdos son respecto al conflicto general que trata el proceso de paz.

**Si estudiamos la distribución de los tipos de acuerdo para cada proceso de paz del top 10 nos encontramos con lo mismo, cada proceso tiene una distribución concreta, con mayoría de un tipo de acuerdo, lo que no tiene nada que ver con la distribución general del tipo de acuerdo.** De nuevo destaca el proceso sirio, que tiene mayoritariamente acuerdos IntraLocal y el proceso del Norte de Irlanda que tiene como mayoría InterIntra, pero también muchos solo de Intra.

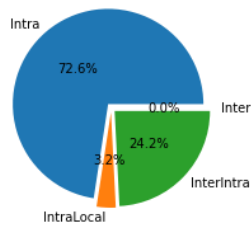
Philippines - Mindanao process



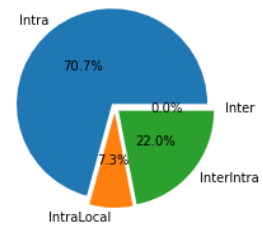
Somalia Peace Process



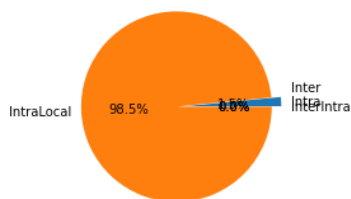
Bosnia peace process



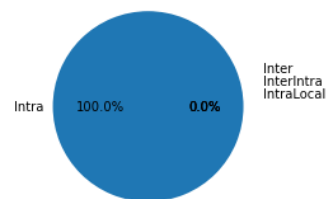
Croatia negotiation process



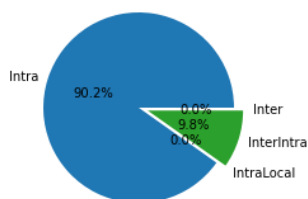
Syrian Local Agreements



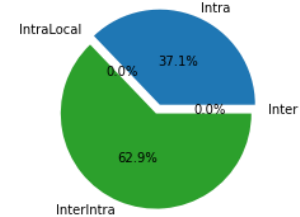
Colombia I - Gaviria



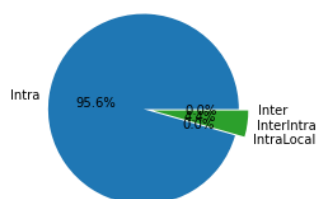
Abkhazia peace process



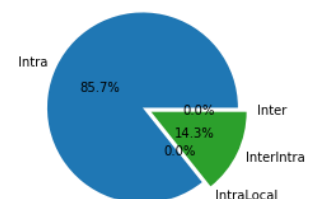
Northern Ireland peace process



Colombia III - Arango

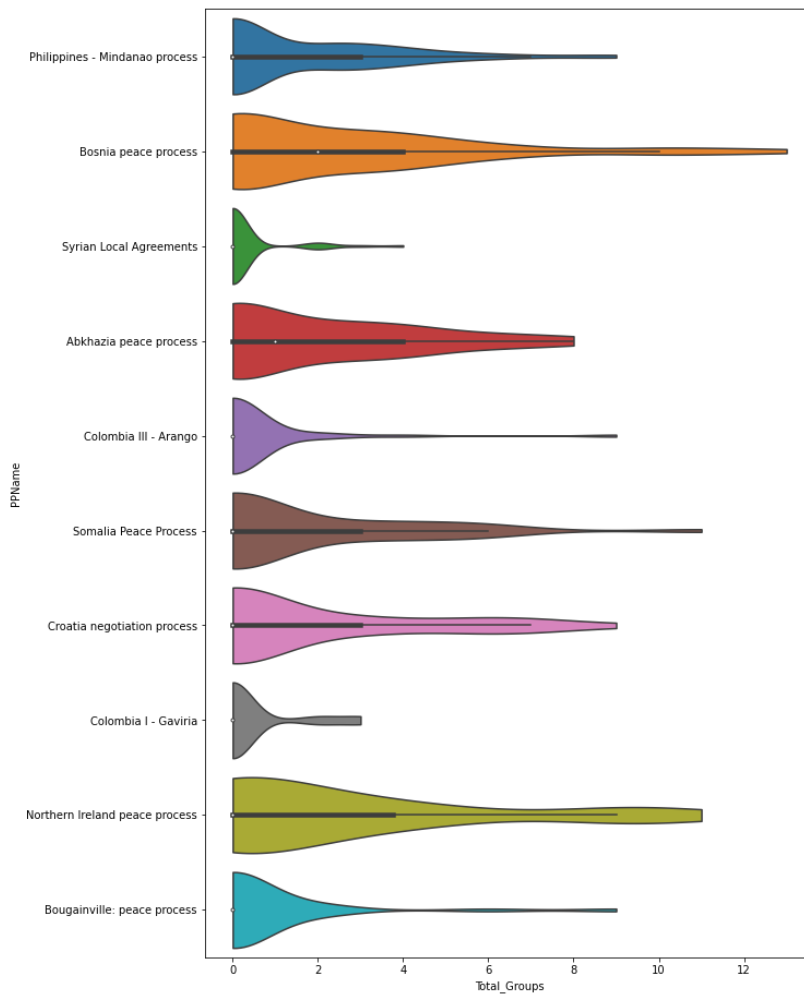


Bougainville: peace process



Esto destaca de nuevo que según los aspectos del conflicto que trata un proceso de paz, los acuerdos internos de un proceso de paz son acordes a este conflicto, similares entre ellos diferentes a los de otros procesos de paz.

Por último, se he encontrado que la distribución de la variable **Total\_Groups**, una variable que indica la importancia del tema grupos en los acuerdos de paz de cada proceso de paz varía en función de este proceso de paz, de nuevo demostrando lo que creíamos, que los acuerdos de paz que pertenecen a un proceso de paz son similares entre ellos, y esto hace que los procesos de paz sean diferentes a otros procesos de paz.



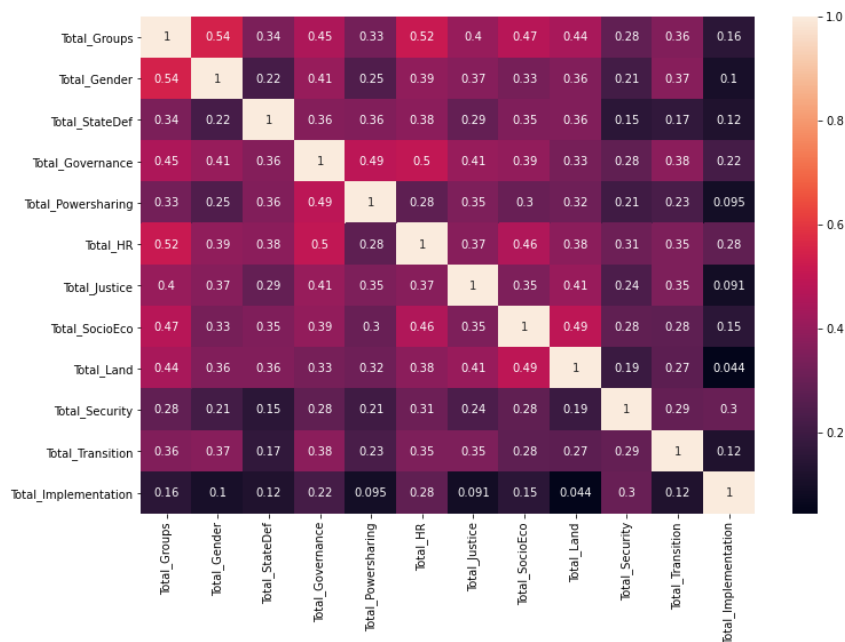
En resumen, se ha encontrado interesante mediante la exploración que los diferentes procesos de paz del top 10 de procesos de paz son diferentes entre ellos respecto a sus características y a los acuerdos que los componen, pero que, dentro de un proceso de paz, los acuerdos que lo conforman son diferentes.

De esta manera, se puede presuponer que un acuerdo de paz está marcado por un conflicto geográfico y trata temas diferentes en épocas diferentes.

Este descubrimiento se desarrollará más en la visualización Dashboard del punto 6.

#### Curiosidades.

Ya se ha visto que los acuerdos internos de un proceso comparten aspectos entre ellos, como el tipo de conflicto, el de acuerdo o incluso lo que se trata un tema como el de grupos. **Se ha estudiado también si entre los diferentes acuerdos de paz hay correlación entre los temas que se tratan más en un acuerdo de paz.** Para esto, se ha obtenido la correlación de las diferentes puntuaciones de cada tema (variables creadas en la transformación) para todos los acuerdos.

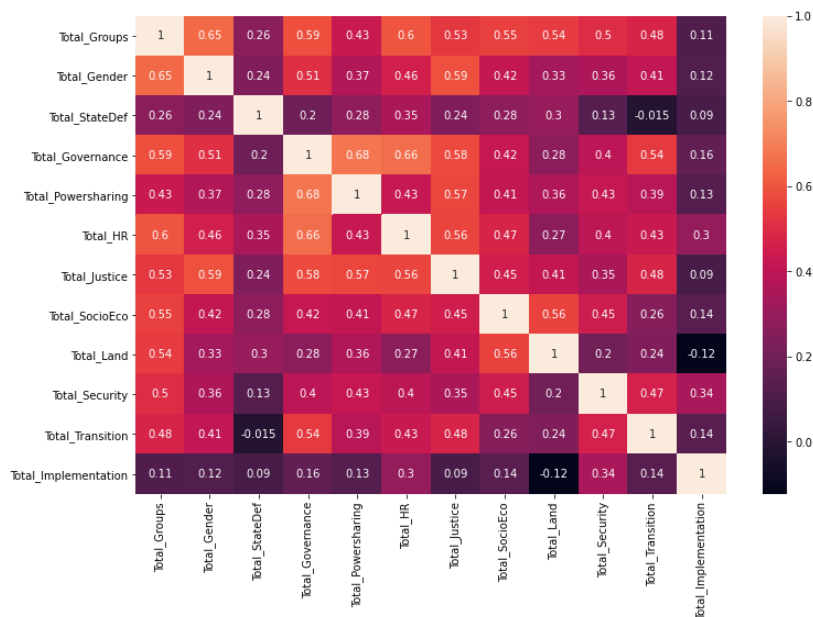


**Vemos como algunos temas de acuerdos tienen mayor correlación con otros.** Vemos que en los acuerdos que se tratan los grupos, se tratan normalmente también los géneros, y los derechos humanos dando una sensación de que hay acuerdos muy sociales/humanísticos.

Por otra parte vemos como hay correlación entre acuerdos de Governance con Powersharing y Governance con HR, pero hay muy poca correlación en los acuerdos entre HR y Powersharing, todo esto podría indicar que cuando se trata el tema de gobierno, se puede hacer desde una visión social, o desde una visión de control del poder total.

También destaca que hay correlación entre los acuerdos con alto interés socioeconómico y aquellos con alto interés de tierra, indicando que estos pueden ser otro tipo de acuerdos que se dan en los procesos de paz.

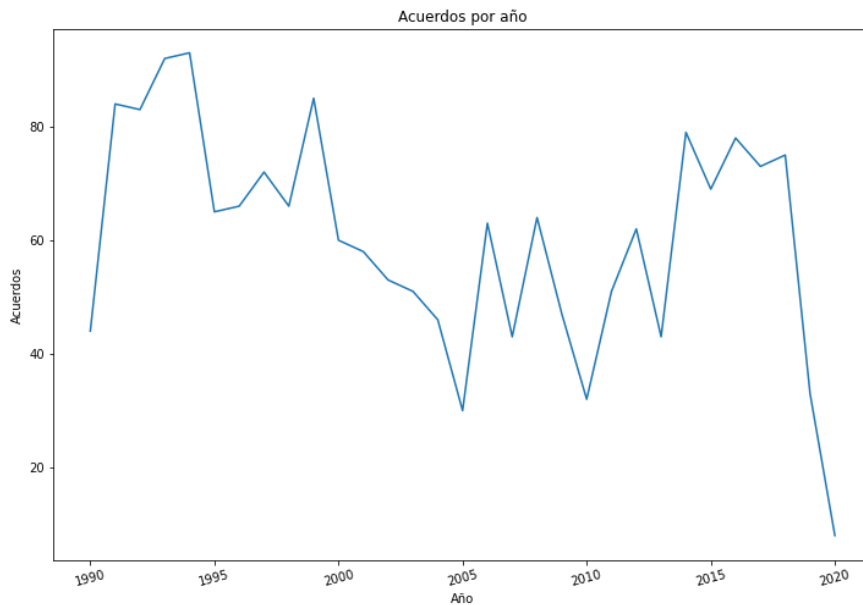
Como ya hemos visto, hay similitudes entre los acuerdos de paz internos de los procesos de paz top 10, a continuación, **se observa que la correlación de la media de los temas que se tratan en un acuerdo entre los diferentes procesos de paz se mantiene**, es decir vemos de nuevo que generalmente en un proceso cuando crece el tratamiento de aspectos de grupos sociales, crece el de género y el de derechos humanos principalmente. Vemos también el resto de las correlaciones mencionadas hasta ahora.



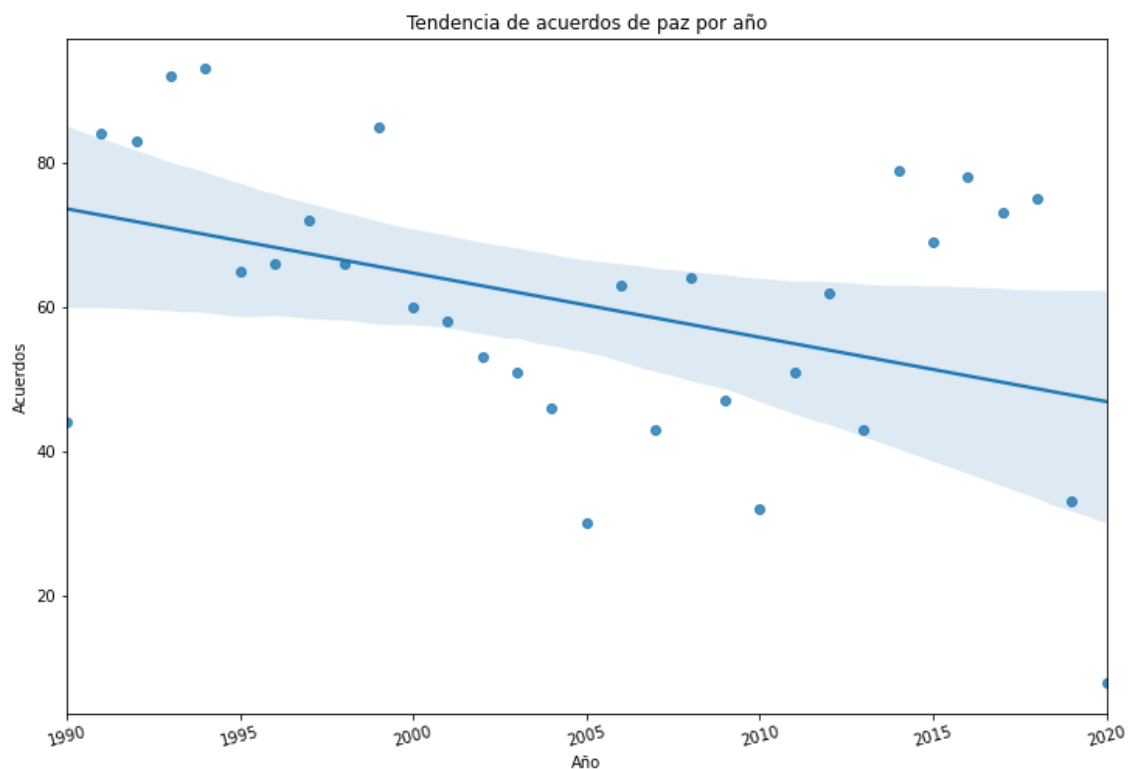
Otro aspecto interesante que se ha encontrado en el dataset, ha sido al estudiar anteriormente la distribución de la variable Total\_Groups entre los diferentes procesos de paz del top 10. Antes de obtener la conclusión, se ha visto que hay outliers en la distribución de la variable para los acuerdos de paz. **Es decir, que, aunque de manera general todos los procesos tienen acuerdos de paz con valores similares para tratar un tema como los grupos, también hay acuerdos que tienen números de la variable muy altos en muchos procesos, lo que significa que hay acuerdos destinados a tratar los grupos de personas en procesos de paz, los cuales tratan mucho este tema.**

Como última curiosidad se ha estudiado en la exploración de los datos la línea temporal de acuerdos de paz durante estos 30 años, podemos ver como en la década de los 90 hubo muchos acuerdos de paz, estos probablemente debidos a la partición de Yugoslavia, que ya hemos visto que es el país con más acuerdos.

A continuación, vemos que durante los 2000 y hasta 2014 hay una bajada de acuerdos de paz por año, sin embargo desde 2015 hasta 2019 ha habido un aumento a casi un nivel de acuerdos de paz de los 90. Por tanto, no queda clara la tendencia.



Para encontrar la tendencia se ha realizado un regplot, que nos permite ver si los acuerdos por año siguen una tendencia negativa o positiva, vemos que es ligeramente negativa, lo que indica que los acuerdos por año decrecen a lo largo del tiempo, ya sea porque hay menos guerras o porque se firman más acuerdos.



5. Los datos, procedimiento y herramientas. Explicad cómo lo habéis descubierto: con qué herramientas, con qué operaciones.

¿Qué proceso has seguido en la pregunta 4?

El proceso seguido ya se puede intuir al leer la pregunta 4. **Partiendo del estudio de los datos y los tipos inicial realizado en la pregunta 3, en la pregunta 4 en resumen, se han obtenido evidencias simples de los datos para comprender inicialmente el conjunto de datos. A continuación, se ha explorado la diferencia que hay entre los procesos de paz top 10 por el número de acuerdos de paz, identificando que estos procesos son diferentes, por tanto, proponiendo como aspecto interesante esto para desarrollar la visualización en la tarea 6. Por último, se ha completado la exploración viendo la correlación que hay entre los temas que se tratan en los acuerdos y la temporalidad de los acuerdos.**

**En la búsqueda de evidencias,** inicialmente, se ha estudiado primero los aspectos básicos de los datos. En concreto se ha visto el número de acuerdos por país y se ha graficado, a continuación, se ha visto el número de acuerdos por proceso de paz y se ha graficado. Para ver el número de acuerdos simplemente se agrupan los datos por la variable y se cuentan, ya sea por país o por proceso de paz. La gráfica estática la he realizado con gráficos de columnas.

A continuación, he comprendido que tres de las variables categóricas más importantes y características de los acuerdos de paz son el tipo de conflicto que resuelven, el tipo de acuerdo que definen y en qué fase del proceso de paz se realiza. Por tanto, siguiendo el mismo procedimiento anterior, he calculado el número de acuerdos por categoría de cada y lo he graficado mediante gráficos de barras.

A partir de todo esto, he podido extraer las evidencias de los datos, que interpreto como aspectos sencillos de encontrar en los datos, hechos evidentes de estos.

**Durante la exploración para encontrar un aspecto interesante para fundamentar la visualización de la tarea 6,** he pensado lo que ya se ha explicado en la tarea 4, que es posible que cada proceso de paz se caracterice por los acuerdos que lo componen, que en realidad vienen caracterizados por el conflicto general a solucionar. De esta manera, los acuerdos de un proceso de paz serán similares entre ellos y diferentes respecto a los de otros procesos, haciendo que un proceso de paz sea diferente a otros procesos de paz, o similares si tratan conflictos similares.

Para esto, primero he limitado el dataset de acuerdos de paz a aquellos que pertenecen a los procesos de paz del top 10. Después, he calculado la distribución del tipo de conflicto y tipo de acuerdo (contando cuando acuerdos por categoría hay y calculando el porcentaje) para cada uno de los procesos del paz top 10, he empezado con estas variables porque son variables fáciles de tratar y que en teoría tienen que cumplir con lo esperado.

Mediante la comparación de un pie chart para cada proceso he visto que es cierto, y que los acuerdos de los procesos son similares entre ellos, haciendo diferentes a algunos procesos de paz respecto al resto. Por otra parte, he estudiado la distribución de la variable que hace referencia a cuanto trata un acuerdo de paz el tema de grupos, generada en la tarea 3.

Primeramente, he observado que hay outliers en las distribuciones de los diferentes procesos, porque todas estas distribuciones están muy estiradas en los violinplot hacia el eje positivo. Por tanto, he detectado los outliers de la variable Total\_Groups mediante el z score y los he eliminado del dataset (aquellos con  $z \geq 3$ ) para hacer la distribución de la variable.

Mediante la distribución sin outliers (graficada con violinplots) he podido ver como la distribución de los procesos de paz se comporta diferente en función del proceso, lo que indica

que hay diferencias entre procesos y entre los acuerdos de paz que componen los diferentes procesos.

**A partir de estas conclusiones de que los procesos de paz son diferentes entre ellos, generaré el Dashboard de la tarea 6.**

**Por último, como curiosidades** he estudiado la correlación entre las variables generadas, que referencian a cuanto se trata un tema general de los acuerdos de paz, como son los grupos, el género, la justicia o la defensa del estado. He calculado la correlación de spearman solo para estas variables, puesto que primero he limitado el dataset a estas variables. A continuación, he graficado la matriz de correlación como un heatmap para extraer las conclusiones explicadas en la tarea 4. Para obtener las correlaciones por proceso de paz usando la media de los acuerdos de los procesos de paz, se ha seguido el mismo procedimiento, pero agregando las variables por su media, después calculando la correlación para las variables según los procesos de paz y por último graficándolas.

Finalmente, he estudiado la temporalidad de los acuerdos de paz y su tendencia. Para esto, he convertido la variable Dat a tipo datetime, he agregado el dataset por el año de Dat y el valor de la agregación es el número de filas, que es lo mismo que el número de acuerdos. De esta manera tengo el número de acuerdos por año.

Lo he graficado y visto la línea de acuerdos por año y estudiado la tendencia de manera gráfica.

**¿Qué software has usado en la PAC?**

**Para realizar el estudio de los datos de la tarea 3, las transformaciones de la tarea 3, y la exploración de los datos de la tarea 4 he utilizado el lenguaje de programación Python 3.7. En concreto, he usado un fichero Jupyter Notebook, a partir del entorno de Anaconda. El fichero Jupyter Notebook se llama “Exploración\_Datos\_PEC2\_VD.ypinb”.**

Primeramente, he utilizado Adobe Reader para leer la documentación de la base de datos.

Para realizar la lectura del dataset inicial, escribir los derivados con las transformaciones, y manejar la mayoría de operaciones del dataset he utilizado el paquete Pandas. En este destaco que para contar el número de acuerdos por categoría en una variable he utilizado la función `value_counts()` que cuenta y ordena automáticamente las categorías. También me ha permitido agrupar los datos al calcular los acuerdos por años mediante un `groupby` y calcular la correlación de los datos de manera muy sencilla mediante `corr()`.

Otros paquetes usados en la PAC para el desarrollo y aspectos concretos han sido `collections`, `numpy` o `math`, principalmente para calcular aspectos puntuales.

Por último, para realizar las visualizaciones estáticas de la exploración he usado `matplotlib` y `seaborn`, dos módulos de Python especializados en visualizaciones y que permiten obtener todo tipo de gráficas como hemos visto hasta ahora.

Finalmente, he usado Tableau Desktop 2021.1 para realizar la tarea 6.

**Herramientas de visualización posibles para crear la visualización de la tarea 6.**

Hay varios tipos de herramientas que podría haber utilizado, como herramientas más sencillas se encuentran herramientas de infografías online como Infogram o DataWrapper, estas herramientas permiten de manera online generar infografías estáticas interesantes y sencillas. Su punto débil y que me han hecho descartarlas es que realmente no permiten realizar grandes



dashboards que permitan explorar los datos como se ha realizado y se espera en la PAC. Además, siguen un modelo Freemium que limita las operaciones a realizar sin pagar, y nosotros necesitamos complejidad en esta PAC.

Otra posibilidad es processing, un lenguaje de programación destinado a las visualizaciones, este lo he descartado rápido por el desconocimiento del lenguaje de programación y la curva de dificultad que puede suponer para realizar la PAC.

Una posibilidad que ofrece una buena interactividad con el dashboard y que permite realizar visualizaciones complejas sería utilizar una librería como D3.js, pero también existen las propias de Matplotlib como plotly.js que el tipo de gráfica ya lo conozco. Estas librerías parten de que hay que conocer javascript y también tienen una curva de dificultad alta, por tanto se han desestimado aunque la libertad de desarrollo que ofrecen es casi infinita, y además se pueden integrar en desarrollo web.

Por último, solo quedan herramientas para el diseño de visualizaciones de datos generadas por empresas como son Power BI, Qlik Sense/View y Tableau. Estas herramientas permiten que a partir de los datos, se utilicen visualizaciones semigeneradas para realizar visualizaciones de datos interesantes e interactivas. La principal ventaja es que son sencillas de empezar con ellas, y que permiten conseguir resultados muy buenos sin tener que aprender lenguajes de programación nuevos. El contrapunto de estas herramientas es que suelen ser de pago.

Me he decidido por Tableau por las ventajas que ofrece la herramienta y porque en el máster se nos ofrece la licencia.

## 6. Visualización sobre los datos. Un dashboard o un conjunto de visualizaciones sobre los datasets escogidos.

La visualización producida mediante Tableau Desktop se encuentra en el link: [https://public.tableau.com/views/Comparativatemporalgeograficaytematicadelos10procesosdepazconmsacuerdos/Dashboard1?:language=es&:display\\_count=y&publish=yes&:origin=viz\\_share\\_link](https://public.tableau.com/views/Comparativatemporalgeograficaytematicadelos10procesosdepazconmsacuerdos/Dashboard1?:language=es&:display_count=y&publish=yes&:origin=viz_share_link)

También se encuentra en los ficheros entregados en la PAC bajo el nombre de “Comparativa temporal, geográfica y temática de los 10 procesos de paz con más acuerdos.twbx”

**En este dashboard se parte de la información interesante obtenida en la tarea 4. En él busco visualizar diferencias en la localización, en la temporalidad y la temática entre el top 10 de los procesos de paz del dataset. Tal y como se ha realizado en el análisis exploratorio, pero con un poco más de profundidad en algunos aspectos y con visualizaciones más complejas y que siguen buenas prácticas.**

**Es importante mencionar que el dashboard cuenta con interactividad por parte del usuario, pudiendo hacer Overview, Zoom, Filter y Details on Demand, según se definen en The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations de Shneiderman.**

**Inicialmente el dashboard hace un pequeño análisis general de los acuerdos de paz, mostrando el número de acuerdos de paz que hay y el número de procesos de paz. A continuación, se muestra los acuerdos por año de una manera muy similar a como se ha mostrado en el análisis**

exploratorio, pero con un área y no solo una línea, lo que se considera una buena práctica porque marca más la atención de quién ve la visualización ya que le da sentido de completitud. También se muestra el número de acuerdos por país/región, pero mediante un mapa, donde los países con más acuerdos tienen un tono más oscuro. De esta manera la visualización permite que el usuario pueda comparar aquellos países con más acuerdos por país debido al tono de color y también fijarse en un país que le interese, aunque no sea un país top en acuerdos, esto es gracias a que los datos se representan sobre una superficie el elemento de más alto nivel de los elementos que facilitan la visualización según Blackwell en Visual Representation.

**A partir de aquí, se explica que los procesos de paz comprenden varios acuerdos de paz, y que en teoría los procesos de paz contienen diferencias entre ellos y sus acuerdos son similares internamente. Por tanto, se estudian el top 10 de los procesos de paz de manera específica.**

**Primero se muestra un gráfico de burbujas con los procesos de paz siendo cada uno una burbuja.** Solo están coloreados aquellos que pertenecen al top 10, cada uno con un color que nos permite identificar el proceso de paz en todas las gráficas. Además, cada bola tiene un tamaño de acuerdo con el número de acuerdos de paz que engloba. Así el usuario que le interese conocer el número de acuerdos puede pasar el ratón sobre la bola de un proceso y conocerlo, pero todos los usuarios pueden comprender las dimensiones de acuerdos por proceso para los procesos con más acuerdos del dataset, porque se entiende que una bola mayor es porque tiene más acuerdos.

**Al lado, hay un gráfico ridgeline, que permite comparar la línea temporal de cada proceso de paz viendo así como algunos se han desarrollado a lo largo de muchos años, mientras que otros ocurrieron principalmente en los 90 y otros son mucho más recientes.** El usuario puede si quiere situarse sobre una de las líneas temporales y ver el número de acuerdos para ese proceso de paz en ese año, pero lo interesante es comprender como los procesos de paz tienen diferentes temporalidades.

**La siguiente gráfica es un mapa de procesos de paz, que nos permite comparar la localización geográfica de los diferentes procesos.** En el mapa hay representados con el mismo color aquellos países o regiones que pertenecen a un mismo proceso de paz, esto es un principio gráfico que indica que aquellos elementos similares (en este caso los países con igual color) pertenecen a un mismo grupo (los procesos de paz), el principio se conoce como principio de similitud, y lo definió Stephen Few en Data Visualization for Human Perception. De esta manera el usuario puede visualizar que se desarrollan procesos de paz en todo el mundo.

**Para estudiar las diferencias temáticas entre los diferentes acuerdos, se comienza con una visualización de dos heatmap, uno para cada variable tipo de conflicto y tipo de acuerdo, que muestran la distribución de acuerdos por proceso para estas variables.** A pesar de que aparece el número en cada casilla, de nuevo se observa como interesante que el tono más fuerte representa un mayor número de acuerdos, para que el visualizador no tenga que centrarse en el número sino quiere y solo comprender que tipo de acuerdo o conflicto destaca en cada proceso de paz. **Estos dos heatmap, son una mejor manera de visualizar los gráficos de tarta que se han realizado en la exploración para identificar igual la distribución de acuerdos y conflictos en cada proceso, puesto que el color permite diferenciar mejor que los ángulos de un gráfico de tarta.**

**Por último en la comparación temática, se ha realizado un parallel coord plot, este permite comparar la importancia relativa de cada tema en los acuerdos de paz, mostrando así si un proceso destaca tratando un tema o dejando de tratar otro, y permitiendo comparar la**

**temática diversa de cada proceso.** Este parallel coord plot es útil porque permite que el visualizador investigue y pueda sacar sus propias conclusiones complejas y lógicas, como la naturaleza temática de un proceso, gracias a que sigue las ideas de Few en Data Visualization for Human Perception, sobre todo la de facilitar la comparación entre cantidades. Cada línea está pintada del color del proceso de paz asociado en los mapas, el ridge line y el gráfico de bolas, para que el usuario pueda interpretarlo fácil.

**Respecto a la interacción,** en el centro de estas gráficas hay una leyenda interactiva que permite que el usuario seleccione un proceso de paz, y las representaciones de este en las gráficas se destacan respecto al resto de procesos, facilitando la comprensión por parte del usuario.

**Para entrar todavía en más conceptos específicos de los datos, y en concreto de los procesos de paz, he realizado una sección dedicada a comparar dos procesos de paz.** Para cada proceso de paz se puede visualizar el mapa centrado en los países que participan en el proceso de paz, esta vez con cada país/región participante de un color diferente para que el usuario pueda identificarlos como separados. Además, se ha realizado un treemap para cada proceso de paz para que el usuario pueda visualizar en cuantos acuerdos de paz ha participado cada país, de nuevo, se utiliza el color para remarcar mayor cantidad de acuerdos de paz por país, aunque en el tree map ya se observa esto mediante el tamaño del cuadrado. Es importante mencionar el aspecto interactivo que tiene este último apartado que permite que se seleccione (filter) mediante un desplegable cada uno de los dos procesos de paz a comparar.

Por último, hay una visualización que muestra la media de valor de los diferentes grupos que se presentan en los acuerdos de un proceso de paz, para ambos procesos, permitiendo así la comparación entre los procesos y ver en que grupos de personas destaca más cada proceso y la comparación entre los procesos. Cada variable tiene el mismo color entre los diferentes procesos para facilitar la comparación de los resultados de cada proceso.