

Visualización de Datos – PEC 2

Carlos Giner Baixauli

1. Título de las visualizaciones que presentan el dataset escogido.

El título seleccionado es “Conflictos y acuerdos de paz en el mundo”. No obstante, el dashboard contiene 5 visualizaciones que muestran la etapa del acuerdo por país, el estado del acuerdo por país, el nivel del acuerdo por tipo de conflicto, el estado del acuerdo por tipo de conflicto y el estado del acuerdo por región.

2. Descripción corta del documento y de los que se presenta.

This document introduces an analysis of the dataset PA-X Peace Agreements, a peace agreements database created and managed by the University of Edinburgh. The dataset is available at www.peaceagreements.org and the current version was released on February 19, 2018.

We will present the interaction between some of the most relevant variables in the dataset in order to visualize their relationship and acquire new knowledge. The dashboard will include 5 visualizations: Stage by country, Agreement status by country, Agreement conflict level by conflict nature, Agreement status by conflict nature and Agreement status by region.

3. Qué he aprendido explorando los datos: tipo, estructura, curiosidades.

El dataset contiene 1789 observaciones y 12 variables. Estas variables son las siguientes:

- *X.U.FEFF.AgreementId* es una variable numérica que identifica el acuerdo de paz.
- *Name* es una variable cualitativa que consiste en el nombre del acuerdo de paz.
- *Region* es una variable cualitativa que identifica la región del acuerdo de paz. Contiene 6 niveles (Europe and Eurasia, Cross-regional, Middle East and North Africa, Africa (excl MENA), Asia and Pacific y Americas).
- *Country* es una variable cualitativa que identifica el país del acuerdo de paz. Contiene 169 niveles.
- *Peace.Process* es una variable numérica que codifica los procesos de paz. Contiene 152 niveles.
- *Peace.Process.Name* es una variable cualitativa que identifica el nombre de los procesos de paz codificados en la variable *Peace.Process*. Contiene 152 niveles.
- *Stage* es una variable cualitativa que identifica la etapa del proceso de paz. Contiene 7 niveles: SubPar (Marco-sustantivo parcial), Imp (Implementación / Renegociación), Pre (Pre-negociación en proceso), SubComp (Marco-sustantivo comprehensivo), Cea (Alto el fuego), Ren (Renovación) y Oth (Otros).

- *Signed.Date* es una variable numérica de tipo fecha que identifica la fecha en la que fue firmado el acuerdo.
- *Agreement.Conflict.Level* es una variable cualitativa que indica el nivel del acuerdo. Contiene 4 niveles: Inter (Entre estados), InterIntra (Entre estados y dentro de un estado), Intra (Dentro de un estado) e IntraLocal (Dentro de un estado, a nivel local).
- *Conflict.Nature* es una variable cualitativa que identifica el tipo de conflicto. Contiene 5 niveles (Government, Government/territory, Inter-group, Territory y Other).
- *Agreement.Status* es una variable cualitativa que indica el estado final del acuerdo. Contiene 4 niveles (Agreement with subsequent status, Multiparty signed/agreed, Status nuclear y Unilateral document).
- *Agreement.Text* es una variable cualitativa que contiene el texto del acuerdo.

Para obtener información visual del dataset, las variables más interesantes son las cualitativas estructuradas en niveles. Estas variables agrupan los datos en categorías, de manera que contando las observaciones por grupo podemos conocer su distribución. Además, si cruzamos estas variables podemos obtener nueva información.

4. Qué habéis descubierto: evidencias, tendencias, outliers, etc.

De las variables discretas citadas en el apartado anterior, obtenemos los siguientes conteos por nivel:

Region

Africa (excl MENA)	529
Americas	195
Asia and Pacific	376
Cross-regional	34
Europe and Eurasia	407
Middle East and North Africa	248

Agreement.Conflict.Level

Inter	72
InterIntra	234
Intra	1293
IntraLocal	190

Stage

Cea	369
Imp	325
Oth	9
Pre	505
Ren	40
SubComp	106
SubPar	435

Conflict.Nature

Government	662
Government/territory	813
Inter-group	206
Other	1
Territory	107

Agreement.Status

Agreement with subsequent status	7
Multiparty signed/agreed	1652
Status unclear	10
Unilateral document	120

De los resultados anteriores y de las visualizaciones presentadas en el dashboard obtenemos algunas conclusiones, entre las que destacan las siguientes:

- El estado del acuerdo “Multiparty signed/agreed” es, con diferencia, el más frecuente tanto a nivel global como a nivel de región.
- La naturaleza del conflicto implica al gobierno en la mayoría de los casos.
- Los acuerdos alcanzados se corresponden con conflictos a nivel interno de los estados en la mayoría de los casos.

Con las variables descritas puede ser interesante conocer la distribución geográfica de los acuerdos con comparaciones como la etapa del acuerdo por país, así como el estado del acuerdo por país y por región. También podemos obtener información valiosa de tipo causa-efecto si comparamos la naturaleza del conflicto con el nivel del conflicto el acuerdo y con el estado del acuerdo.

Finalmente resaltamos las tendencias que hemos indicado en este apartado, así como la presencia de un outlier, el valor “Other” en Conflict.Nature y en el cruce de esta variable con otras variables cualitativas discretas.

5. Explicar cómo lo habéis descubierto: con qué herramientas, cálculos, con qué operaciones.

Para la realización del apartado anterior se ha procedido a contar los casos en cada categoría de las variables seleccionadas para observar su distribución, su tendencia y la presencia de outliers. También se han cruzado algunas de estas variables para obtener grupos de casos en función de combinaciones de variables.

Para las operaciones con una variable se ha utilizado el lenguaje de programación R, con el que se pueden realizar los cálculos mencionados de forma muy sencilla. No obstante, para la combinación de variables ha sido preferible el uso de Tableau, ya que permite la visualización de patrones y características de una manera mucho más rápida y sencilla.

El campo de Data Science presenta una gran variedad de herramientas para el análisis como los lenguajes R y Python. Estos lenguajes también disponen de librerías de visualización como ggplot2, matplotlib y seaborn, no obstante, existen herramientas más específicas y desarrolladas para este fin como Tableau, que es una herramienta de BI enfocada en la visualización y que permite crear dashboards. Tengo previsto conocer también otras herramientas como QlikView, similar en algunos aspectos a Tableau, y D3.js, una librería para producir infografías en JavaScript.

6. El Dashboard o un conjunto de visualizaciones sobre los datos.

El dashboard está disponible en el siguiente enlace:

https://public.tableau.com/profile/carlos.giner#!/vizhome/Libro1_15741944240630/ConflictsandPeaceAgreementsaroundtheWorld