

Visualización de datos

PEC 2

NADIA SÁNCHEZ POZO

1. **Título:** Acuerdos de paz a nivel mundial.

2. **Descripción:**

En el presente trabajo se realiza un análisis exploratorio del conjunto de datos recuperado de Peace Agreements, una base de datos de documentos sobre acuerdos de paz a nivel mundial desde 1990.

El conjunto de datos consta de diferentes acuerdos, desglosados por país / entidad, región, tipo de conflicto, tipo de acuerdo y etapa del acuerdo. En el caso de tipo de conflicto el tipo de conflicto que aborda el acuerdo es: Gobierno, territorio, Gobierno / territorio, Otro. 'Gobierno' se refiere a disputas ideológicas o políticas, y 'Territorio' es puramente disputas territoriales. Todas las disputas intraestatales sobre el territorio están codificadas como 'Gobierno / territorio'. Los atributos son cualitativos y cuantitativos relacionados con política, derecho, seguridad, reformas, reconstrucción socioeconómica, justicia transicional, desarrollo e implementación. Los datos se pueden analizar en conjunto o de forma separada por grupos. En el siguiente enlace podemos encontrar una descripción de cada uno de los atributos. https://www.peaceagreements.org/files/PA_X_codebook_Version2_June_2019.pdf

En este análisis exploratorio vamos a describir los datos encontrados, para conocer los tipos de datos que forman el dataset, identificar características tales como: valores atípicos o outliers, tendencias, y patrones en los datos.

3. **Los datos, presentación:**

El conjunto de datos está formado por 1789 observaciones de 266 variables, las cuales tienen parámetros numéricos, categóricos y también tipos texto natural. En este conjunto de datos tenemos algunos metadatos sobre cada columna. Existen variables binarias que toman valores de 0 ó 1, como por ejemplo las variables “Groups”, “State definition” y sus subgrupos.

Dentro de las principales variables tenemos las siguientes:

1. Con: País / Entidad
2. Contp: tipo de conflicto
3. PP: Número del proceso de paz
4. PP_name: Nombre del proceso de paz
5. Reg: Región
6. AgtId: ID del acuerdo
7. Agt: Nombre del acuerdo
8. Dat: Fecha firmada

9. Status: definición y estado del acuerdo
10. Lgt: longitud del acuerdo (páginas)
11. Agtp: tipo de acuerdo / conflicto
12. Stage: etapa de acuerdo
13. StageSub: subcodificación de etapa
14. Part
15. Loc1GWNO

A partir de estas variables podemos empezar a plantearnos algunas preguntas sobre los datos, por ejemplo, ¿Qué tendencias siguen las variables?, con las cuales se pueden plantear preguntas concretas relacionadas a los acuerdos como: ¿Cuál es el país con mayor número de acuerdos?, ¿Tipo de conflicto predomina en américa?, ¿En qué año se suscitaron más conflictos?, etc.

4. Los datos, exploración:

En este apartado se realiza un análisis descriptivo de los datos, con el objetivo de encontrar relaciones, distribuciones, valores atípicos, etc.

En el conjunto de datos podemos apreciar la existencia de valores perdidos, aproximadamente más del 90% de variables son binarias, además una variable puede estar dividida en varios subgrupos formando otras variables. Por lo que se observan altas correlaciones superiores al 80%. Además, existen otras variables que su único valor es igual a 0 como es el caso de las siguientes variables: GeLgbti, GeLgbtiPos, GOthOth, GeLgbtiNeg, GMigOth, GAgeOth, HrNiOth. La variable lmsrc, toma un único valor igual a "1".

5. Los datos procedimientos y herramientas.

En primer lugar, para garantizar la calidad del dato, es necesario realizar una limpieza del conjunto de datos. Para esto utilizamos la librería pandas de Python, eliminando aquellas variables que tengan más del 5% de valores NaN, eliminamos las variables: ThrdPart', 'OthAgr', 'Loc2ISO', 'Loc2GWNO', 'UcdpCon', 'UcdpAgr', 'PamAgr', 'CowWar'.

Luego definimos una matriz de correlación para obtener el índice de variables con correlación superior al 80% ya que este tipo de variables altamente correlacionadas aportan información duplicada que está contenida en otras variables. Por lo tanto, eliminamos las siguientes variables 'N_characters', 'GChSubs', 'GDisSubs', 'GAgeSubs', 'GMigSubs', 'GRaSubs', 'GReAntid', 'GReSubs', 'GIndSubs', 'GRefSubs', 'GeMeNu', 'EpsRes', 'HrNiNe', 'CeProv', 'TjAmPro', 'TjRMa'.

Este proceso se especifica en nsanchezpoPEC2.ipynb

El siguiente paso es cargar nuestros datos a una herramienta más robusta como tableau, en la cual encontramos nuestros datos divididos entre medidas y dimensiones, esta funcionalidad es de gran utilidad ya que separa las operaciones que se pueden realizar con cada atributo en función de su naturaleza.

6. Vista sobre los datos.

El objetivo de este dashboard es tener una mejor perspectiva de los datos. Por lo cual se ha optado por presentar tres visualizaciones dependientes en donde podemos filtrar los datos por tipo de conflicto en la parte superior tenemos un Heatmap de acuerdos por país y conflicto, y en la parte inferior tenemos un diagrama de área donde se representa el número de acuerdos por año y etapa del acuerdo (stage), en el lado izquierdo tenemos una gráfica de puntos de región y estado del acuerdo donde representamos algunas métricas como:

- Cor: corrupción, toma el valor de 1 (generales), 2 (proceso concreto) o 3 (detallado) si el acuerdo de paz incluye alguna mención de medidas para abordar la corrupción.
- CitGen: Ciudadano general, variable binaria, valor de 1 si el acuerdo incluye referencias generales a ciudadanos que no sean cuando usan la palabra en lugar de "todas las personas". Si no hay tales disposiciones el valor en esta variable es 0
- DevHum: asistencia humanitaria. Esta es una variable binaria, que indica si el acuerdo contiene disposiciones para asistencia humanitaria (1, 0 de lo contrario).
- DevSoc: desarrollo socioeconómico. Esta es una variable binaria, que indica si el acuerdo contiene disposiciones para desarrollo en general (1, 0 de lo contrario).
- HrCit: Ciudadanía, representa cualquier referencia de acuerdo a la ciudadanía, en relación con cualquier cambio.
- GAge: Esta variable explica cualquier mención de personas mayores, estipulaciones basadas en la edad o similares, toma valores de 1, 2 o 3.
- JusCr: JusCr: justicia penal y derecho de emergencia. Variable binaria, toma el valor de 1 si el acuerdo de paz incluye disposiciones relacionadas con la justicia penal, caso contrario es 0.

URL de la visualización:
https://public.tableau.com/views/nsanchezpoPEC2/Dashboard1?:retry=yes&:display_count=y&:origin=viz_share_link

7. Referencias

Bell, Christine, Sanja Badanjak, Robert Forster, Astrid Jamar, Sean Molloy, Kevin McNicholl, Kathryn Nash, Jan Pospisil, and Laura Wise (2019). PA-X Peace Agreements Database and Dataset, Version 2. www.peaceagreements.or