

1. Contexto. Explicar en qué contexto se ha recolectado la información. Explicar por qué el sitio web escogido proporciona esta información.

Después de hacer diferentes búsquedas de páginas que pudieran contener datos interesantes y comentarlo entre nosotros, decidimos que no queríamos generar un dataset típico con precios de productos, pisos, etc. Así que buscando fuentes con otro tipo de datos, encontramos la web de Knoema y pensamos que podría permitir hacer comparativas muy interesante entre países dado que todos los indicadores han sido extraídos con los mismos criterios.

2. Definir un título para el dataset. Escoger un título que sea descriptivo.

Principales indicadores macroeconómicos y demográficos por país.

3. Descripción del dataset. Desarrollar una descripción breve del conjunto de datos que se ha extraído (es necesario que esta descripción tenga sentido con el título escogido).

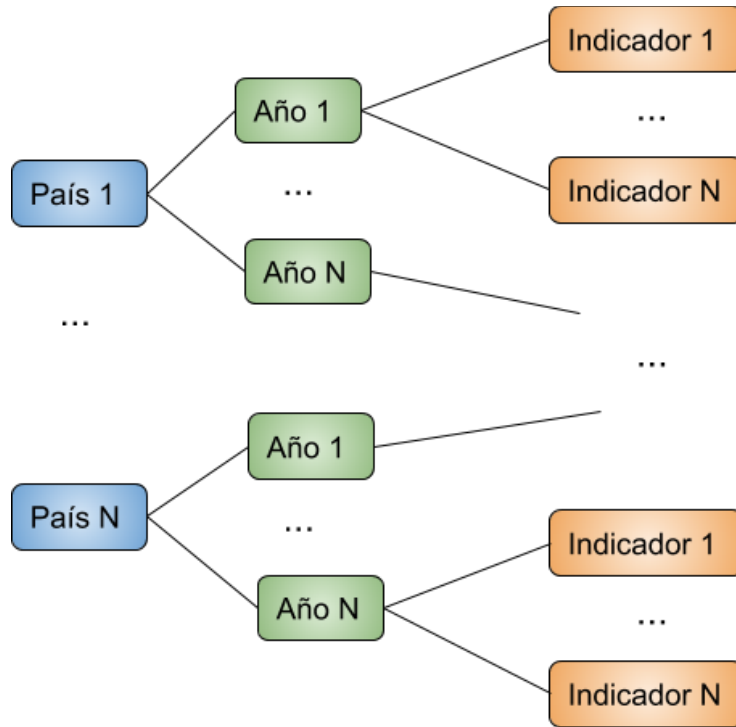
El dataset está compuesto de un conjunto de indicadores económicos y demográficos que expresan claramente la evolución de los países a lo largo de la última década.

4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.

La siguiente imagen representa el dataset en su conjunto:



Mientras que esta otra representa la estructura concreta de la información:



5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se han recogido.

El dataset consta de los siguientes campos:

- País: Nombre del país.
- Año: Año del dato.
- PIB: Producto interior bruto. Expresado en billones de dólares. Campo numérico
- Inflación del IPC: Subida del índice de precios al consumo. Expresado en porcentaje.
- Tasa de desempleo: Tasa de población sin trabajo. Expresado en porcentaje.
- Importación de bienes y servicios: Compra de bienes o servicios provenientes de otro país u otro mercado distinto del propio. Expresado en dólares.
- Exportación de bienes y servicios: Venta de bienes o servicios a otro país. Expresado en dólares.
- Población: Residentes legales en un país. Expresado en personas.
- Densidad de la población: Número de personas por metro cuadrado. Expresado en personas.
- Tasa de natalidad: Cantidad de nacimientos en un país por cada 1000 habitantes. Expresado en nacimientos.
- Tasa de mortalidad: Cantidad de muertes en un país por cada 1000 habitantes. Expresado en muertes.

- Esperanza de vida al nacer: Número de años que en promedio esperaríamos vivir una persona si durante toda su vida estuviera sujeta a las condiciones de mortalidad por edad observadas en el período de estudio. Expresado en años.
- Relación empleo/población: Proporción de la población en edad laboral que está trabajando. Expresada en porcentaje.
- Terrenos agrícolas (km²): Terrenos dedicados a la agricultura o usados como pasto. Expresado en km².
- Zona forestal: Áreas arborizadas, fuera de zonas agrícolas y centros urbanos. Expresado en porcentaje.
- Mortalidad infantil: Número de niños que mueren antes de llegar al año de vida por cada 1.000 habitantes. Expresado en muertes.
- Gasto en salud, % del PIB: Gasto en salud expresado en función del PIB. Expresado en porcentaje.
- Gasto en salud per cápita: Gasto en salud por persona. Expresado en dólares.
- Gasto militar, US\$ actual: Gasto en defensa. Expresado en millones de dólares.
- Gasto militar, % del PIB: Gasto en defensa en función del PIB. Expresado en porcentaje.
- Gasto público en educación: Gasto en educación por el conjunto de las administraciones públicas. Expresado en %.
- Tasa de homicidios: Muertes infringidas por otra persona por cada 100.000 habitantes. Expresado en muertes.
- Ingresos por turismo: Ingresos resultantes de los gastos de los visitantes extranjeros. Expresado en dólares.
- Gasto en investigación y desarrollo, % del PIB: Gasto en investigación y desarrollo en función del PIB. Expresado en porcentaje.
- Índice de Desarrollo Humano: Índice compuesto que mide el rendimiento promedio en tres dimensiones básicas del desarrollo humano: una vida larga y saludable, conocimiento y un nivel de vida digno, mejor cuanto más cercano a 1. Expresado en unidades.
- Felicidad: Índice publicado por la Naciones Unidas calculado según diferentes factores como el PIB per cápita. Expresado en unidades.
- Usuarios de internet, % de la población: Usuarios de internet en función del total de la población. Expresado en porcentaje.

Los datos de acceso libre (sin cuenta) hacen referencia a los últimos 12 años como mucho, pero depende de cada indicador (algunos pueden mostrar los 4 últimos, por ejemplo).

En cuanto a cómo se han recogido los datos, hemos diseñado un programa en Python que, a partir de un listado de indicadores, busca en la página principal de Knoema o en cualquiera de sus subpáginas (/topics) la URL correspondiente para cada indicador y país. Después, se descarga la página y parsea el año y el valor guardándolo en una

estructura. Finalmente, el programa recorre esta estructura e imprime los valores en un archivo con formato CSV.

6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si hay).

Los datos están extraídos de [Knoema](#), para indicar a qué se dedican, citaremos como se definen ellos mismo en su WEB: *“Knoema es la fuente más completa de datos de inteligencia empresarial global del mundo. Nuestras herramientas permiten a individuos y organizaciones explorar, visualizar, modelar y presentar sus datos y datos mundiales para facilitar mejores decisiones y resultados.”*

Nosotros hemos accedido a la versión FREE de los datos, pero también tienen las opciones profesional y enterprise, que necesitan usuario para su explotación.

7. Inspiración. Explicar por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.

Como hemos comentado más arriba, este conjunto de datos nos parece muy interesante porque permite comparar diferentes indicadores económicos, demográficos, etc. (hay muchos más indicadores en la web que no hemos extraído en este ejercicio), que pueden ayudar a responder preguntas de tipo “¿cómo han afectado las políticas de austeridad a la evolución del gasto en salud?”, “¿cómo ha evolucionado la esperanza de vida en cada país?”, etc., así como también hacer comparativas entre países, continentes, etc., por ejemplo, “¿cómo es la tasa de natalidad en los países del sur de Europa respecto a la del norte?”.

8. Licencia. Seleccionar una de estas licencias para el dataset resultante y explicar el motivo de su elección:
 - Released Under CC0: Public Domain License
 - Released Under CC BY-NC-SA 4.0 License
 - Released Under CC BY-SA 4.0 License
 - Database released under Open Database License, individual contents under Database Contents License
 - Other (specified above)
 - Unknown License

El uso anónimo de la web no requiere la aceptación de términos y condiciones específicos (sí es obligatorio hacerlo si se crea una cuenta), por tanto, podemos aplicar la licencia que queramos.

En nuestro caso, hemos escogido, CC BY-NC-ND ya que:

- permite copiar y distribuir el contenido (siempre atribuyendo la autoría original)

- no permite el uso comercial (ya que nosotros no hemos generado los datos, no nos parece ético que alguien pueda sacar provecho económico del trabajo de Knoema)
- no permite distribuir derivados (por ejemplo, puedes modificar los datos pero no estás autorizado a distribuirlos).

9. Código. Adjuntar el código con el cual se ha generado el dataset.

Incluído en el archivo comprimido entregado donde se encuentra este documento.

10. Dataset. Presentar el dataset en formato CSV.

Incluído en el archivo comprimido entregado donde se encuentra este documento.
Como separador de campos hemos usado “;”.

Contribuciones	Firma
Búsqueda previa	Gervasio Cuenca, Sabela de la Torre
Redacción de las respuestas	Gervasio Cuenca, Sabela de la Torre
Desarrollo código	Gervasio Cuenca, Sabela de la Torre