

Exploración de datos en repositorios públicos



Adrián Yared Armas de la Nuez

Temas a tratar en la actividad descrita:

Paso 1: Accede a los siguientes sitios web:

- Instituto Nacional de Estadística (INE)
- Eurostat
- Organización Mundial de la Salud (OMS) - Estadísticas de Salud Mundial
- UCI Machine Learning Repository
- Kaggle
-

Paso 2: Explora las secciones de datos disponibles en estos sitios y selecciona una categoría o conjunto de datos que te interese, como estadísticas de salud, datos de la Unión Europea o conjuntos de datos de aprendizaje automático.

Paso 3: Examina los atributos de los conjuntos de datos que elijas, prestando especial atención a la fuente de los datos, fechas de obtención, actualización y validez, así como la metodología de recogida o muestreo.

Paso 4: Anota la información relevante sobre el conjunto de datos que hayas seleccionado.

Paso 5: Si deseas un desafío adicional, busca problemas o desafíos de datos planteados en Kaggle por empresas y organizaciones.

para realizar esta actividad he escogido Eurostat y he decidido analizar el apartado “Unemployment”, es decir, el apartado que se encarga de analizar la tasa de desempleo por sexo, año, edad y país, tanto como la tendencia general de los veintisiete países de la Unión Europea y el área Europea que consta de veinte países que data a partir de 2023.

A través de la observación de los datos y métodos asociados a su recolección y gestión, puedo decir lo siguiente; la obtención de datos depende de cada institución nacional de estadística, en el caso de España el INE (Instituto Nacional de Estadística) el cual en la tramitación de su información tiene tres artículos; Adaptación económica de la población activa en el marco Europeo del primer cuarto de 2021; Resumen de cambios en los económicamente activos Encuesta de Población (EPAS) en 2021 y Media del efecto de los cambios económicos de la población activa a través de (EPAS) en 2021. Esto supone la actualización del método de estudio estadístico de la población activa y la desempleada en el marco de este país concreto, no obstante, el estudio estadístico público general es visible a partir de enero de 1983 como podemos observar en la tabla “Unemployment by sex and age”. A continuación apporto una captura de esta información:

Unemployment by sex and age – monthly data

Online data code	une_rt_m
Source of data	Eurostat
Last data update	07/10/2024 22:00 (1 day ago)
Last structure update	Eurostat 07/10/2024 22:00
Overall data coverage	1983-01 - 2024-09

Resumiendo:

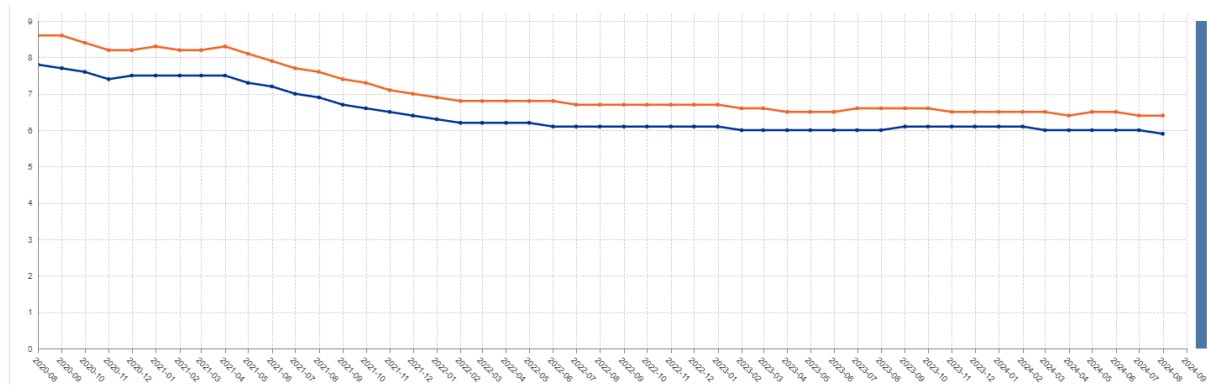
Tiene una validez desde enero de 1983 y su última actualización fue el 07/10/2024 22:00 y se recogen los datos a través de cada entidad nacional de estudio estadístico, pero siguiendo una metodología estandarizada por la Unión Europea, que es la siguiente <https://ec.europa.eu/eurostat/web/lfs/quality> y en cuanto a la diseminación y formato sigue la base de datos <https://ec.europa.eu/eurostat/data/database> .

En cuanto al análisis de datos, podemos observar que el dataset contiene una información estructurada. y sigue una normativa expresa en el metadata relacionado a la base de datos, esta relata la unidad de medida(años, meses, ...), el período (meses de referencia), el mandato del marco legal, el tratado de confidencialidad de los datos, la calidad de la gestión y otros datos como la accesibilidad y claridad de la información.

Sobre mi elección concreta de información, puedo destacar la siguiente información; Consiste en el análisis del desempleo y mercado laboral, enfocado en el análisis por sexo (hombres y mujeres) y edad, con actualizaciones mensuales y principalmente mide la tasa de desempleo y población activa. Todo esto con el fin de analizar el mercado laboral, elaborar políticas públicas, estudios socioeconómicos, y para que los gobiernos e

instituciones monitoreen y gestionen el desempleo de forma eficaz. A través de una información libre, clara y descargable en diferentes formatos.

Y a un nivel más general, hasta la fecha de hoy hemos visto que el estudio está teniendo un resultado positivo, ya que la tasa de desempleabilidad está en descenso, como podemos observar en el análisis de los datos generales:



Además de esto, he buscado un problema de la comunidad en Kaggle, este consiste en predecir la asistencia a través de una información histórica y demográfica de los habitantes, se nos aportan los tipos de datos, acceso a los datos, las reglas, el período de la duración del desafío, el método de evaluación y al código inicial.

El desafío es la siguiente página:

<https://www.kaggle.com/competitions/desafio-comunidad/data>

Páginas utilizadas y relacionadas a la información analizada:

https://ec.europa.eu/eurostat/cache/metadata/EN/employ_esgrs_es.htm

https://www.ine.es/en/daco/daco42/daco4211/epa0121_en.pdf

https://www.ine.es/en/inebaseDYN/epa30308/docs/epa_cambios2021_en.pdf

https://www.ine.es/en/inebaseDYN/epa30308/docs/medida_efecto_epa_cambios2021_en.pdf

https://ec.europa.eu/eurostat/databrowser/view/UNE_RT_M_custom_13219125/default/table?lang=en

https://ec.europa.eu/eurostat/cache/metadata/EN/employ_esms.htm