



Universitat
Oberta
de Catalunya

Tipología y Ciclo de Vida del Dato

Práctica 1 – Web Scraping

Alumno: Rodrigo Baranda Castrillo

Índice de Contenidos

Pregunta 1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.	3
Pregunta 2. Definir un título para el dataset. Elegir un título que sea descriptivo.....	3
Pregunta 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 elegido).....	3
Pregunta 4. Representación gráfica. Presentar una imagen o esquema que identifique el dataset visualmente.....	3
Pregunta 5. Contenido. Explicar los campos que incluye el dataset, el periodo de tiempo de los datos y cómo se ha recogido.....	4
Pregunta 6. Agradecimientos. Presentar al propietario del conjunto de datos. Es necesario incluir citas de investigación o análisis anteriores (si los hay).....	4
Pregunta 7. Inspiración. Explique por qué es interesante este conjunto de datos y qué preguntas se pretenden responder.	4
Pregunta 8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:	4
Pregunta 9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.	5
Pregunta 10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.	5

Pregunta 1. Contexto. Explicar en qué contexto se ha recolectado la información. Explique por qué el sitio web elegido proporciona dicha información.

Se ha decidido analizar los datos macroeconómicos por país y año relativos a la tasa de desempleo, salario medio y salario mínimo con el objetivo de ver cómo han ido evolucionando estos datos en los países a lo largo de los últimos 15 años. Además, se tratará de analizar posibles correlaciones entre Tasa de Desempleo – Salario Mínimo Profesional – Salario Medio.

Además de realizar este análisis descriptivo, un posible alcance de la práctica sería aplicar un algoritmo de clusterización y ver si esto se corresponde con las conclusiones llegadas vía análisis descriptivo; sin embargo, al trabajar con series temporales, quizá esto no sea aplicable.

Webs descargadas:

- <https://datosmacro.expansion.com/paro>
- <https://datosmacro.expansion.com/smi>
- <https://datosmacro.expansion.com/mercado-laboral/salario-medio>

Pregunta 2. Definir un título para el dataset. Elegir un título que sea descriptivo.

El título elegido es “Datos Macroeconómicos [2005-2020]”.

Pregunta 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 elegido).

El dataset contiene los datos macroeconómicos por país desde el año 2005 hasta el año 2020. Se han consolidado los datos relativos a tasa de desempleo, salario mínimo interprofesional y salario medio.

Para crear este dataset, inicialmente se han descargado todos los datos mencionados y se ha realizado una consolidación (join). Al realizar esto, se ha perdido información, dado que hay países para los que puede haber datos de desempleo, pero no de salario mínimo, etc. Por lo tanto, además del dataset principal objeto de análisis, se han descargado los demás datos por si se considera necesario analizarlos más adelante.

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

El dataset cuenta con 441 filas y 6 columnas. En la siguiente imagen se puede ver una captura de la información subida en git.

	Año	País	Tasa Desempleo	Varianza Anual Desempleo	SMI €	Salario Medio
0	2005	España [+]	8,8%	-1,50	598,5 €	20.616€
1	2005	Reino Unido [+]	5,1%	0,40	1.134,7 €	44.361€
2	2005	Francia [+]	9,1%	0,20	1.217,9 €	30.521€
3	2005	Portugal [+]	8,8%	0,50	437,2 €	14.042€
4	2005	Estados Unidos [+]	4,9%	-0,50	655,4 €	30.252€
5	2005	Japón [+]	4,4%	-0,10	826,3 €	36.382€
6	2005	Australia [+]	5,1%	0,00	1.184,7 €	32.213€







Además de este dataset, los otros descargados contienen la siguiente información:

Tabla	Filas	Columnas
Datos Macroeconómicos	441	6
Tasa Desempleo	697	6
Salario Medio	632	4
Salario Mínimo	2.139	6

Como ya se ha mencionado, mantenemos estos datasets por si queremos hacer un análisis más exhaustivo de la distribución de estas variables.

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

Los datos que contiene nuestro dataset son:

-  Año. Año de publicación de los datos. Contenemos información entre 2005 y 2020.
-  País. Países para los cuales tenemos información.
-  Tasa Desempleo. Porcentaje de personas desempleadas en ese país en un año concreto.
-  Varianza Anual Desempleo. Variable que indica cómo ha variado la tasa de desempleo en el último año.
-  SMI (€). Salario Mínimo Interprofesional del país en ese año.
-  Salario Medio (€). Salario Medio del país en ese año.

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

El propietario de los datos es la web Expansión, en su apartado de Datos Macroeconómicos. A su vez, esta web no indica de dónde obtiene los datos.

Ésta página es una web española especializada en materia económica, teniendo una sección de periodismo tradicional y otra sección en la que ponen a disposición del usuario información como la que se ha recogido de datos macroeconómicos, pero también información de datos de bolsa, divisa, materias primas...

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

El objetivo del análisis a realizar es el de descubrir relaciones entre las variables Salario Mínimo – Tasa de Desempleo y Salario Medio. Por ejemplo, poder ver si países con un elevado salario mínimo, en contraposición tienen un bajo salario medio, etc.

Por otro lado, una vez realizado el análisis descriptivo y se haya podido descubrir los tipos de países que hay en función de estas características, me gustaría aplicar un algoritmo de clusterización, y ver si se llegan a los mismos resultados; teniendo en cuenta que se trabaja con series temporales y que esto dificulta el proceso.

Pregunta 8. Licencia. Seleccione una de estas licencias para su dataset y explique el motivo de su selección:

Se ha elegido la licencia Released Under CC BY-SA 4.0 License:

- Deberán seguir con esta licencia las contribuciones a este trabajo: se asegura que se seguirá utilizando una licencia, como mínimo, tan restrictiva como ésta.
- Cuando se hagan cambios sobre el trabajo, habrá que indicar dichos cambios y el autor original: se sigue reconociendo la autoría y qué cosas ha aportado cada uno.

- Se permite comercializar, con lo que podremos conseguir llegar a más cantidad de personas si alguien utiliza nuestro dataset con este fin.

Pregunta 9. Código. Adjuntar el código con el que se ha generado el dataset, preferiblemente en Python o, alternativamente, en R.

El código se encuentra en el enlace:

<https://github.com/rbaranda10/AnalisisMacroeconomico>

Pregunta 10. Dataset. Publicación del dataset en formato CSV en Zenodo (obtención del DOI) con una breve descripción.

El DOI obtenido es:

10.5281/zenodo.4264921

November 9, 2020

Datos Macroeconómicos [2005-2020]

Rodrigo Baranda

Se ha decidido analizar los datos macroeconómicos por país y año relativos a la tasa de desempleo, salario medio y salario mínimo con el objetivo de ver cómo han ido evolucionando estos datos en los países a lo largo de los últimos 15 años. Además, se tratará de analizar posibles correlaciones entre Tasa de Desempleo – Salario Mínimo Profesional – Salario Medio.

Preview

Name	Size	Preview	Download
datos_macro.csv	24.0 kB		

md5:aed8b666653177234581283bef5dfdc6

Citations 0

Show only: ☐ Literature (0) ☐ Dataset (0) ☐ Software (0) ☐ Unknown (0)

☐ Citations to this version

No citations.

Indexed in

OpenAIRE

Publication date:
November 9, 2020

DOI:
DOI: 10.5281/zenodo.4264921

Keyword(s):
Economics Data Science

License (for files):
Creative Commons Attribution 4.0 International

Versions

Version	Date
Version 1.0	Nov 9, 2020

<https://zenodo.org/record/4264921#.X6l5GWhKiHs>