

Instrumento de evaluación del RA1 (ce a, b, c, d, e) y RA2 (ce b, d, e):

A partir de las hojas de cálculo proporcionadas, realiza las siguientes actividades:

Actividad 1: Ingesta de datos. (2 puntos)

- Establece el espacio de trabajo con setwd.
- Utiliza la función read_ods para cargar todos los datos de interés de la hoja de cálculo agua.ods en el dataframe **datosAgua**. Los datos que te interesan están en el rango A7:S121, no cargues cabecera ni pié. Como ayuda se proporciona el manual del paquete readODS.
- Utiliza la función que quieras para cargar los datos del fichero censo.csv en el dataframe **datosCenso**. Convierte a entero los datos de la columna Total.
- Añade los comentarios oportunos para explicar las sentencias utilizadas, prestando atención especialmente a los parámetros empleados en la carga de los datos.

Actividad 2: Filtrado de datos. (3 puntos)

- Filtra los datos del dataframe datosAgua, de forma que solo nos quedemos con los registros etiquetados como "2. Volumen de agua no registrada" y las columnas correspondientes al año 2020. Como resultado obtendremos el dataframe **datosAguaFiltrados** que contendrá una columna con el nombre de la autonomía y el volumen de pérdida, de la forma:

	autonomía	perdidas
1	Total Nacional	1065740
2	01 Andalucía	220806
3	02 Aragón	46071
4	03 Asturias, Principado de	33020
5	04 Balears, Illes	28394
6	05 Canarias	70534
7	06 Cantabria	21277

- Ahora filtra los datos del dataframe datosCenso para extraer solo los datos correspondientes a las comunidades autónomas, sin desglosar por provincias ni sexo (sexo = "Total") y correspondientes al año 2020. Como resultado obtendremos el dataframe **datosCensoFiltrados** que contendrá una columna para la autonomía y otra para la población, de la forma:

	autonomía	poblacion
1	01 Andalucía	8484804
2	02 Aragón	1331938
3	03 Asturias, Principado de	1012117
4	04 Balears, Illes	1183415
5	05 Canarias	2178924
6	06 Cantabria	584708

- Añade los comentarios oportunos para explicar las sentencias utilizadas en el filtrado de los datos, comenta los problemas encontrados y las soluciones adoptadas.

Actividad 3: Repara los datos. (3 puntos)

Ya habrás podido adivinar que nuestra intención es unir los dos dataframes en uno, pero tenemos varios problemas. Aunque ambos tengan el mismo número de filas, estas no se corresponden. Si tomamos datosAguaFiltrados como referencia, a datosCensoFiltrados le falta el Total Nacional y los datos de “Ceuta y Melilla” no se encuentran agrupados sino desglosados. Como si fuera poco, el censo total de población de Melilla en 2020 no se calculó (o ‘alguien’ lo ha borrado maliciosamente). Debemos reparar datosCensoFiltrados antes de unir los dos conjuntos de datos, para ello:

- a) Calcula el total del censo de Melilla en 2020 sumando hombres y mujeres, de la forma más elegante posible y arregla el valor en el dataframe datosCensoFiltrados. El resultado debe ser 86450.
- b) Agrupa los registros de “18 Ceuta” y “19 Melilla” en un único registro “Ceuta y Melilla” cuyo valor será la suma de ambos. Borra los registros originales. (sí, ya lo se, arreglamos Melilla y en seguida lo borramos, caprichos del profesor).
- c) Calcula y agrega el “Total Nacional” al dataframe datosCensoFiltrados.
- d) Puedes realizar los cálculos, borrado y agregado de registros como prefieras. Lo más elegante posible. Documenta y explica cada paso. Comenta los problemas encontrados, las soluciones adoptadas.

Tras realizar las tres modificaciones, el dataframe quedará de la forma:

▲	autonomía	▼	poblacion	▼
1	01 Andalucía		8484804	
2	02 Aragón		1331938	
3	03 Asturias, Principado de		1012117	
4	04 Balears, Illes		1183415	
5	05 Canarias		2178924	
6	06 Cantabria		584708	
7	07 Castilla y León		2385223	
8	08 Castilla - La Mancha		2052505	
9	09 Cataluña		7749896	
10	10 Comunitat Valenciana		5067911	
11	11 Extremadura		1061636	
12	12 Galicia		2698177	
13	13 Madrid, Comunidad de		6726640	
14	14 Murcia, Región de		1518279	
15	15 Navarra, Comunidad Foral de		662032	
16	16 País Vasco		2212628	
17	17 Rioja, La		319444	
18	Ceuta y Melilla		170521	
19	Total Nacional		47400798	

No te preocupes porque el Total Nacional esté el último, es indiferente.

Actividad 4: Construye el conjunto de datos. (2 puntos)

- a) Une ambos dataframes en uno nuevo llamado conjuntoDatos.

	autonomía	perdidas	poblacion
1	01 Andalucía	220806	8484804
2	02 Aragón	46071	1331938
3	03 Asturias, Principado de	33020	1012117
4	04 Balears, Illes	28394	1183415
5	05 Canarias	70534	2178924
6	06 Cantabria	21277	584708
7	07 Castilla y León	62225	2385223
8	08 Castilla - La Mancha	65831	2052505
9	09 Cataluña	144526	7749896
10	10 Comunitat Valenciana	114118	5067911
11	11 Extremadura	30607	1061636
12	12 Galicia	64988	2698177
13	13 Madrid, Comunidad de	64700	6726640
14	14 Murcia, Región de	23934	1518279
15	15 Navarra, Comunidad Foral de	15929	662032
16	16 País Vasco	32974	2212628
17	17 Rioja, La	16749	319444
18	Ceuta y Melilla	9057	170521
19	Total Nacional	1065740	47400798

Esta actividad puede realizarse saltando la actividad tres, pero debes buscar la forma de que no aparezcan los registros correspondientes a las filas 18 y 19 de la captura anterior.

- b) Es el momento de sacar al Analista de Datos que tienes dentro. Calcula dos magnitudes que nos informen de como de eficiente es el suministro de agua en cada comunidad, teniendo en cuenta que las pérdidas en el suministro están expresadas en miles de metros cúbicos. Coméntalas y explica su cálculo.