

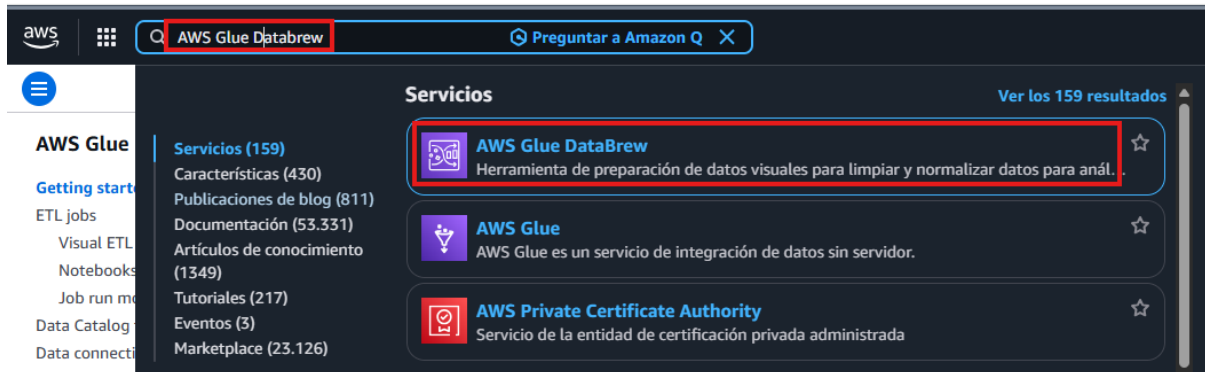
PR_08.3 Dani Gayol Rodríguez

PR_08.3 Dani Gayol Rodríguez.....	1
Apartado A.....	1
1.) Realiza el ejemplo de Proyecto Databrew que se muestra en el tema	
UT_08.4_AWS DataBrew entre las páginas 1 y 7.	2
2,) Muestra también el linaje resultante de los datos.	16

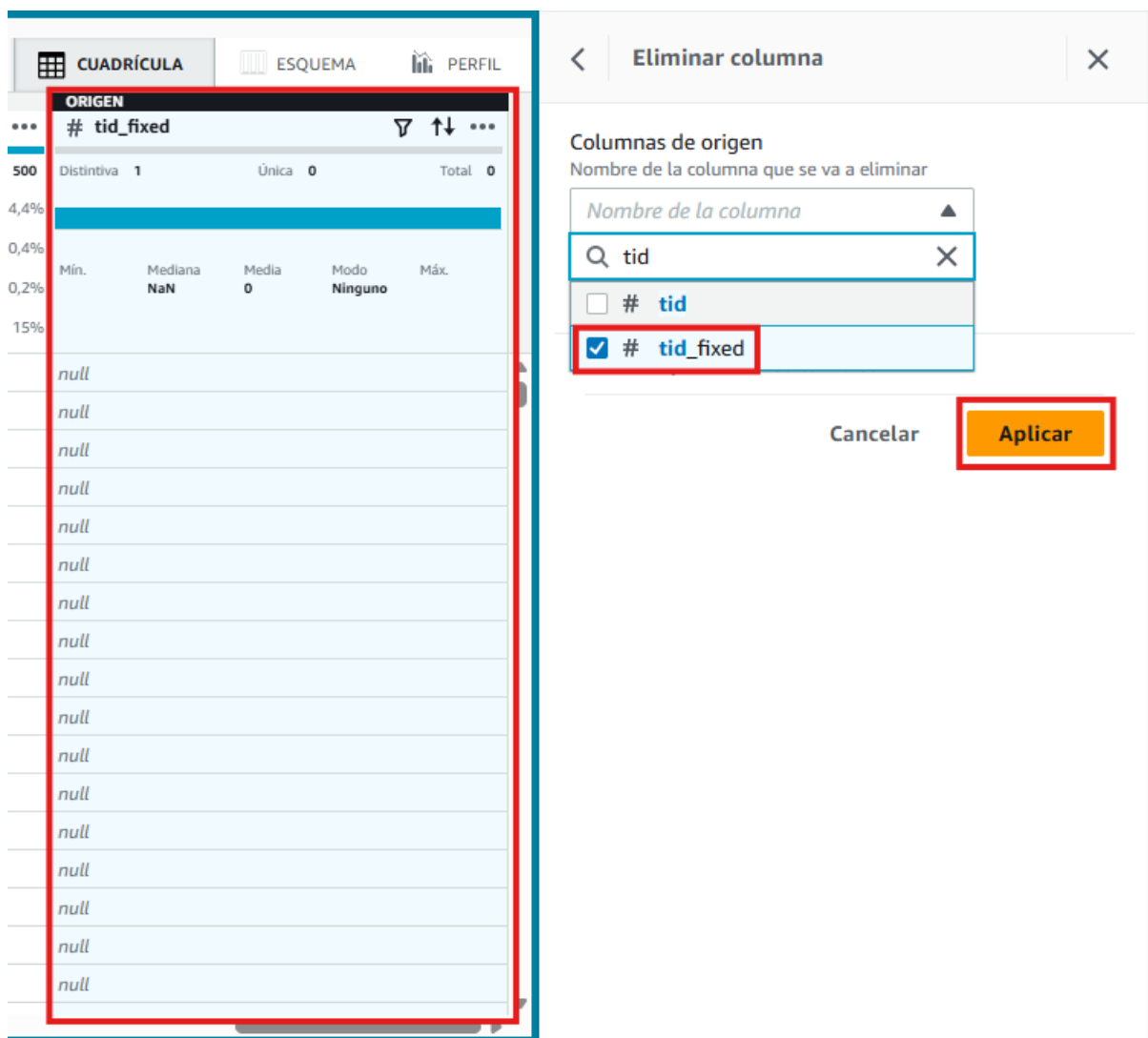
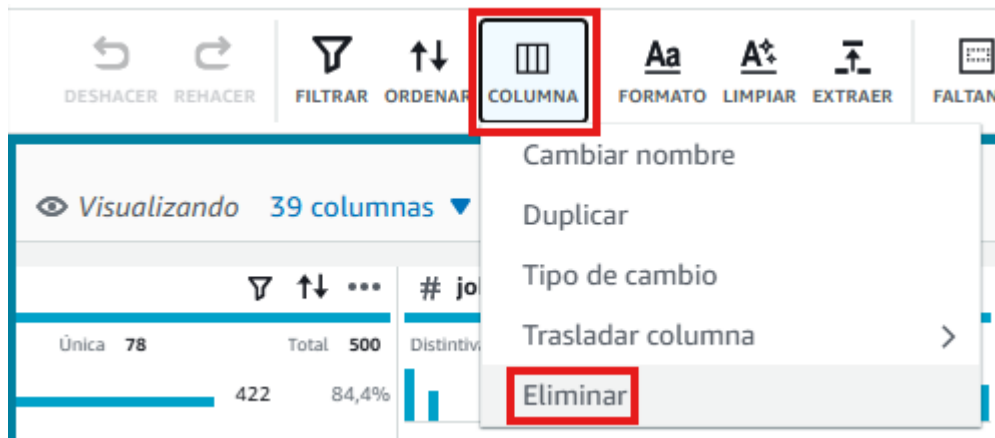
Apartado A

1.) Realiza el ejemplo de Proyecto Databrew que se muestra en el tema UT_08.4_AWS DataBrew entre las páginas 1 y 7.

El primer paso va a ser crear el proyecto de muestra con los datos de CHEMBL



Ahora vamos a seguir con los pasos que nos indica la práctica, el primer paso es eliminar la columna “tid_fixed”



Receta (1)

Sample recipe - 1

Versión de trabajo

Publicar

Más

El paso Eliminar columna se ha aplicado correctamente.

Pasos aplicados (1) | [Borrar todo](#)

1. Eliminar columna `tid_fixed`

Ahora, vamos a filtrar los datos:

DESACER

REHACER

FILTRAR

ORDENAR

COLUMNA

FORMATO

LIMPIAR

EXTRAER

FALTANTE

NO ES VÁLIDO

DUPLICADOS

Visualizando 3

ABC curated_by

Distintiva 3

Autocuration

Intermediate

Expert

Filtros condicionados aplicados

No hay filtros aplicados

Agregar nuevo filtro

Valores faltantes

Es válido

Por condición

count

Única 13

Total 470

ABC ass

Distintiva

ORIGINAL

	14	2,8%	1	2
Autocuration			1	
Autocuration			1	
Autocuration			3	
Expert			23	
Autocuration			48	
Autocuration			1	
Autocuration			22	
Autocuration			1	
Expert			1	
Autocuration			48	

Es exactamente

No es

Contiene

No contiene

Comienza por

Termina con

Menor que

Menor o igual que

Mayor que

Mayor o igual que

Está entre

<

Valores de filtro

×

Nombre de la columna que se va a filtrar

curated_by

▼

Condición de filtro

Es exactamente

▼

☒ Ingresar valor personalizado

☐ Ingresar un valor de RegEx

Introducir un valor de filtro

Autocuration

×

Find

▣

Valores distintivos (1)

Mostrar 1 valor seleccionado

☒

Autocuration

390

78%

☐

Intermediate

96

19%

☐

Expert

14

2%

👁

Vista previa de los cambios

Valores faltantes



Columna de origen

Nombre de la columna con valores faltantes

assay_organism



Acción de valor faltante

Acción que se debe realizar en los valores que faltan

- ☐ Eliminar filas con valores faltantes
- ☐ Rellenar con un valor vacío
- ☐ Rellenar con nulo
- ☐ Rellenar con el último valor válido
- ☐ Rellenar con el valor más frecuente
- ☒ Rellenar con valor personalizado
- ☐ Rellenar con agregación numérica

Valor personalizado

Unknown

Aplicar transformación a

- ☒ Todas las filas (500 filas)
La transformación se aplicará a todas las filas del conjunto de datos
- ☐ Filas filtradas: 0 filtros aplicados(500/500 filas)
La transformación se aplicará a las filas filtradas en la cuadrícula

ABC assay_organism			
Distintiva	81	Única	55
		Total	500
Homo sapiens	203	40,6%	
Rattus norvegicus	62	12,4%	
Mus musculus	57	11,4%	
Todos los demás valores	178	35,6%	
Homo sapiens			
Mus musculus			
Homo sapiens			
Homo sapiens			
Mus musculus			
Unknown			
Homo sapiens			
Rattus norvegicus			
Mus musculus			
Homo sapiens			
Unknown			
Rattus norvegicus			
Unknown			
Homo sapiens			
Haemophilus influenzae			
Homo sapiens			
Oryctolagus cuniculus			

Receta (2)

Sample recipe - 1

Versión de trabajo

Publicar

Más

Pasos aplicados (2) | [Borrar todo](#)

1. Eliminar columna `tid_fixed`

2. Rellenar valores faltantes con `Unknown` en `assay_or`
`ganism`

Ahora, vamos a crear una columna nueva:

DIVIDIR

FUSIONAR

CREAR

Σ

FUNCIONES

CONDICIONES

ANIDAR-DESANIDAR

DINAMIZAR: GRUPO

UNIR

COMBINACIÓN

TEXTO E:

Q

Buscar

Funciones matemáticas

>

Funciones de agregación

>

Funciones de texto

>

Funciones de fecha

>

Funciones de ventana

>

Funciones web

>

Otras funciones

>

Expresión personalizada

>

FECHA

HORA

DATETIME

DATEADD

DATEDIFF

DATEFORMAT

MES

MONTHNAME

AÑO

TRIMESTRE

DÍA

WEEKDAY

WEEKNUMBER

HORA

MINUTO

SEGUNDO

MILISEGUNDOS

nica

55

Total

85

415

83%

13

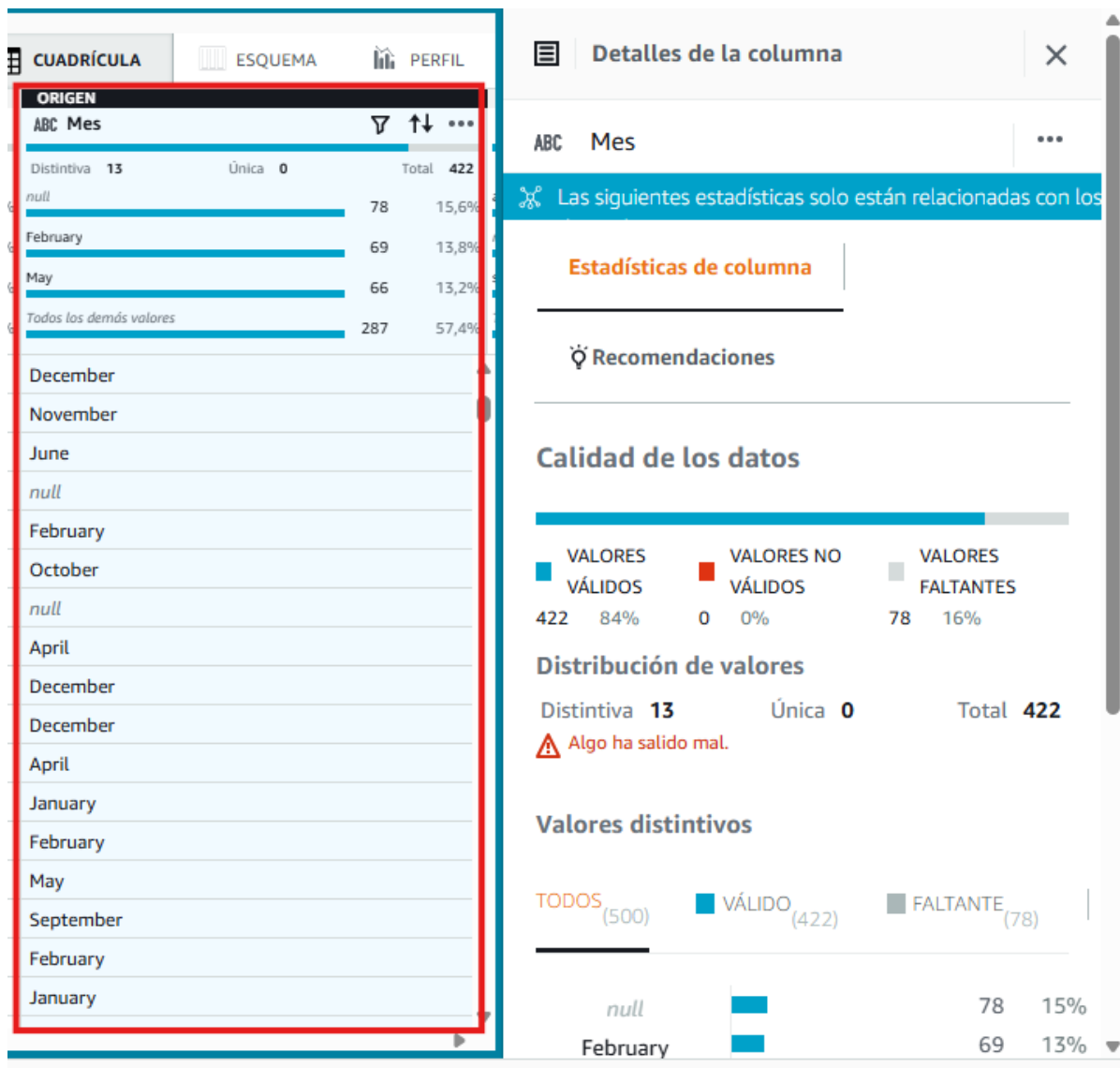
2,6%

8

1,6%

64

12,8%



El siguiente paso es publicar la receta, para ello le tenemos que dar al botón de publicar y ya nos aparecerá en el menú de la izquierda de “Recetas”

Receta (3)

Sample recipe - 1

Versión de trabajo

Publicar

Más

Pasos aplicados (3) | [Borrar todo](#)

1. Eliminar columna `tid_fixed`

2. Rellenar valores faltantes con `Unknown` en `assay_or`
`ganism`

3. Crear columna `Mes uso de` de Función `dateTime MONTH`
`_NAME`

CONJUNTOS

PROYECTOS

RECETAS

REGLAS DE DQ

TRABAJOS

NOVEDADES

DataBrew > Recetas

¿Qué es una receta?

Recetas (1)

Descargar como YAML

Descargar como JSON

Crear trabajo con esta receta

Acciones

Cargar receta

Buscar recetas

Publicado

Nombre de la receta

Descripción de la versión

Proyectos asociados

Fecha de publicación

Publicado por

Etiquetas

Sample recipe - 1

Versión publicada 1.0

-

Sample project - 1

hace unos segundos

26 de enero de 2026, 9:47:59 am

voclabs

-

Una vez dentro de la “receta”, le daremos al botón para crear un trabajo:

Crear trabajo con esta receta

Acciones

ABRIR PROYECTO

Pasos de la receta

Linaje de datos

Detalles de la receta

Nombre de la receta

Sample recipe - 1

Descripción de la versión

-

Proyecto asociado

Sample project - 1

Trabajos asociados

-

Fecha de publicación

hace un minuto por user3935192=Daniel_Gayol_Rodríguez

26 de enero de 2026, 9:47:59 am

Pasos de la receta (3)

1. Eliminar columna `tid_fixed`

2. Rellenar valores faltantes con `Unknown` en `assay_organism`

3. Crear columna `Mes uso de` de Función `dateTime MONTH_NAME`

Crear trabajo

Detalles del trabajo

Nombre del trabajo

Identificador del trabajo

DataBrew-danigayol

El nombre del trabajo debe contener entre 1 y 240 caracteres. Los caracteres válidos son alfanuméricos (A-Z, a-z, 0-9), guiones (-), puntos (.) y espacios.

Tipo de trabajo

Tipo de trabajo que se va a ejecutar en el conjunto de datos



Crear un trabajo de receta

Ejecuta las transformaciones de la receta asociada en la población del conjunto de datos asociado.



Crear un trabajo de perfil

Genera un resumen y estadísticas que le dan la forma de los datos.

Entrada del trabajo

El conjunto de datos de entrada para el trabajo y la receta que se le aplicará.

Ejecutar en



Conjunto de datos

Ejecute el trabajo en un conjunto de datos DataBrew existente o nuevo.



Proyecto

Ejecute el trabajo en un proyecto sin trabajo asociado.

Elegir conjunto de datos

chembl-27

Explorar conjuntos de datos

Conectar nuevo conjunto de datos

Seleccionar una receta

Sample recipe - 1

Versión de la receta

Versión 1.0

Examinar recetas

Salida 1

Configuración

Remover

Salida a

Ubicación de la salida

Amazon S3

Tipo de archivo

Formato de salida

CSV

Delimitador

Separador CSV

Coma (,)

Compresión

Tipos disponibles

None

Cuenta de AWS del propietario del bucket de S3

Cuenta de AWS actual

043356869404

Otra cuenta de AWS

Ubicación de S3

El formato es s3://bucket/folder/

s3://databrew-practica-danigayol/



Explorar

Resumen de configuración

Almacenamiento de salida de archivos

Crear una nueva carpeta para cada ejecución de flujo de trabajo

Salida de archivo

Autogenerate files

División personalizada por valores de columna

Desactivado

Vista previa de la ruta de salida

s3://databrew-practica-danigayol/DataBrew-danigayol_26ene.2026_timestamp_part00000.csv

Salida 2

Configuración

Remover

Salida a

Ubicación de la salida

Amazon S3

Tipo de archivo

Formato de salida

PARQUET

Delimitador

Separador CSV

Coma (,)

Compresión

Tipos disponibles

None

Cuenta de AWS del propietario del bucket de S3

Cuenta de AWS actual

043356869404

Otra cuenta de AWS

Ubicación de S3

El formato es s3://bucket/folder/

s3://databrew-practica-danigayol/



Explorar

Resumen de configuración

Almacenamiento de salida de archivos

Crear una nueva carpeta para cada ejecución de flujo de trabajo

Salida de archivo

Autogenerate files

División personalizada por valores de columna

Desactivado

Vista previa de la ruta de salida

s3://databrew-practica-danigayol/DataBrew-danigayol_26ene.2026_timestamp_part00000.parquet

Permisos

DataBrew needs permission to connect to data on your behalf. Use an IAM role with the [política necesaria](#) attached.

Nombre del rol

Elige el rol que tiene acceso para conectarse a los datos. Actualice para ver las últimas actualizaciones.

LabRole

Cancelar

Crear trabajo

Crear y ejecutar trabajo

Se ha creado el trabajo de receta "DataBrew-danigayol".

DataBrew > Trabajos > DataBrew-danigayol

DataBrew-danigayol

Conjunto de datos: chembl-27

Receta: Sample recipe - 1

Ejecutar trabajo

Acciones

Historial de la ejecución del trabajo

Detalles del trabajo

Linaje de datos

Última ejecución de trabajo 2 minutos, no hay ejecuciones de trabajos programadas

Historial de la ejecución del trabajo

Detener la ejecución del trabajo

Acciones

Buscar por ID de ejecución de trabajo

Mostrar todo

ID de ejecución de trabajo

Estado de la última ejecución del trabajo

Tiempo de ejecución

Salida

Resumen

Iniciado por

Iniciado el

Finalizado el

DataBrew-danigayol_2026-01-26-10:06:52

Realizado con éxito

1 minuto, 30 segundos

2 salidas

user:9935192:Daniel_Gayol_Rodríguez

hace 2 minutos
26 de enero de 2026, 10:06:52 am

hace unos segundos
26 de enero de 2026, 10:06:52 am

Finalmente, nos vamos a “S3” y ya nos saldrán las carpetas:

databrew-practica-danigayol

Información

Objetos

Metadatos

Propiedades

Permisos

Métricas

Administración

Puntos de acceso

Objetos (2)

Copiar URI de S3

Copiar URL

Descargar

Abrir

Eliminar

Acciones

Crear carpeta

Cargar

Los objetos son las entidades fundamentales que se almacenan en Amazon S3. Puede utilizar el [inventario de Amazon S3](#) para obtener una lista de todos los objetos de su bucket. Para que otras personas obtengan acceso a sus objetos, tendrá que concederles permisos de forma explícita. [Más información](#)

Buscar objetos por prefijo

< 1 >

Nombre

Tipo

Última modificación

Tamaño

Clase de almacenamiento

DataBrew-danigayol_26Jan2026_1769418489718/

Carpeta

-

-

-

DataBrew-danigayol_26Jan2026_1769418502926/

Carpeta

-

-

-

DataBrew-danigayol_26Jan2026_1769418489718_part00000.csv

Información

Propiedades

Permisos

Versiones

Información general sobre el objeto

Propietario

d1efcee244946868bbfc7ed543f4012c35bbbed396dc1f3889c27be6adc76c1e3

Región de AWS

EE.UU. Este (Norte de Virginia) us-east-1

Última modificación

26 Jan 2026 10:08:23 AM CET

Tamaño

3.2 MB

Tipo

CSV

Clave

DataBrew-danigayol_26Jan2026_1769418489718/DataBrew-danigayol_26Jan2026_1769418489718_part00000.csv

DataBrew-danigayol_26Jan2026_1769418502926_part00000.parquet [Inform](#)

Propiedades

Permisos

Versiones

Información general sobre el objeto

Propietario

d1efcee244946868bbfc7ed543f4012c35bbbed396dc1f3889c27be6adc76c1e3

Región de AWS

EE.UU. Este (Norte de Virginia) us-east-1

Última modificación

26 Jan 2026 10:08:26 AM CET


Tamaño

2.2 MB

Tipo

parquet

Clave

 DataBrew-danigayol_26Jan2026_1769418502926/DataBrew-danigayol_26Jan2026_1769418502926_part00000.parquet

2.) Muestra también el linaje resultante de los datos.

Para mostrar el Linaje de Datos, nos vamos al “job” ejecutado anteriormente y entramos en el menú de “Linaje de Datos”:

Linaje

Registros de CloudTrail

