Para utilizar Glue podríamos hacer lo siguiente si tenemos grandes cantidades de datos en donde tenemos muchos archivos con el mismo contenido:

1. Se debe asociar una SQS al Bucket q para generar una notificación de eventos cada vez que se creen nuevos archivos en el Bucket
2. La cola SQS se asocia a una Lambda para activar la lambda cada vez que se creen los nuevos eventos (archivos creados)
3. Realizar las transformaciones necesarias para crear Dataframes usando pandas, ya que son pequeños set de datos, igualmente y crear una función adicional para invocar Glue.
4. Se guarda la información enriquecida en Bucket en stage
5. Luego Glue se activa debido a la creación de nuevos archivos enriquecidos en el Bucket de stage, para que Glue empiece a leer los archivos, primero se debe validar en Glue que la cola este vacía, es decir que la lamba no tenga mas eventos
6. Se crean las transformaciones y validaciones necesarias para crear un solo archivo que unifique todos los dataset enriquecidos, esta tarea debe ser ejecutada diariamente
7. Una vez el archivo se encuentra unificado, puede ser consumido por el siguiente proceso para su ingesta o transformación.

Para realizar un proceso CI/ CD para subir la imagen de una lambda, se puede hacer lo siguiente:

* Crear el pipeline en donde se valide y cree lo siguiente:

Implementar la version de Python

Ejecutar las librerías de Requirements.txt

Ejecutar el repositorio git donde se encuentra la dockerizacion de la imagen lambda

Implementar la limpieza de flujo