

## Informe de la calidad de los datos para las bases BDUA

**Población Base de Datos Única de Afiliados (BDUA) del régimen contributivo:** En esta base contiene la población activa afiliada al régimen contributivo del sistema general de seguridad social en salud, incluyendo cotizantes y beneficiarios.

**Población Base de Datos Única de Afiliados (BDUA) del régimen subsidiado:** En esta base incluye a los afiliados que no pagan directamente o aportan poco, es decir, la población focalizada que recibe subsidios de salud. La información es reportada por EPS subsidiadas y entidades territoriales.

Para el análisis se unieron ambos regímenes con el fin de obtener la población total del sistema de salud colombiano. A partir de esta integración, se identificará:

- El tipo de dato de cada variable.
- Tipos de inconsistencias presentes, como duplicados y valores nulos.
- Posibles estrategias de manejo para cada escenario.

El análisis se enfocará en los meses de **septiembre y octubre**, teniendo en cuenta que la actualización de las bases, tanto para el régimen contributivo como subsidiado, se realiza de manera mensual.

### Unión de las bases para el mes de septiembre

Para el mes de septiembre, se unieron las bases del sistema de salud y social correspondientes a los **regímenes contributivo y subsidiado**. Esta integración permite contar con la población total afiliada al sistema de seguridad social en salud en Colombia, facilitando la identificación de tipos de datos, inconsistencias (como duplicados y valores nulos) y posibles estrategias de limpieza y manejo de la información.

VARIABLE	TIPO DE DATO	DESCRIPCION	
hash_sha256_num	Texto / cadena	Identificador único del afiliado	58265933216991557.
Género	Categórica	Sexo del afiliado	Femenino, Masculino
Grupo etario	Categórica / Rango	Edad en rangos por años	(1 a 5), (15 a 19), (50 a 55),ETC..
Código de la entidad	Texto / Código	Identificador de la EPS	EPS541,ETC..
Nombre de la entidad	Texto	Nombre completo de la EPS	NUEVA EPS S.A., EPS SURAMERICANA S.A EMSSANAR S.A.S.,ETC...
Régimen	Categórica	Tipo de régimen de afiliación	Contributivo, Subsidiado
Tipo de afiliado	Categórica	Rol del afiliado dentro del régimen	BENEFICIARIO, CABEZA DE FAMILIA

VARIABLE	TIPO DE DATO	DESCRIPCION	
Estado del afiliado	Categórica	Estado actual de la afiliación	Activo, Protección Laboral C
Condición del beneficiario	Categórica	Condición especial del afiliado	NO APLICA, VICTIMAS DEL CONFLICTO ARMADO INTERNO, ETC...
Zona de Afiliación	Categórica	Tipo de zona donde reside	Urbana, Rural, Urbana-Cabecera Municipal
Departamento	Categórica	Departamento de residencia	CHOCÓ, ANTIOQUIA,ETC...
Municipio	Categórica	Municipio de residencia	ACANDI, SABANETA,ETC...
Nivel del Sisbén	Categórica / Entero	Nivel de puntaje del Sisbén	N,NAN,1,2,O,C,A,B,3,4
Grupo poblacional del afiliado	Categórica	Clasificación poblacional	POBLACION CON SISBEN, VÍCTIMAS DEL CONFLICTO ARMADO INTERNO, ETC...
Cantidad de registros	Numérica (entero)	Número de registros o unidades asociadas al afiliado (frecuencia / duplicados)	1, 2, 15
Fecha de actualización	Fecha / Date Timé	Fecha y hora de la actualización del registro	2025-09-06 22:32:26
id fecha	Texto / Código	Identificador asociado a la fecha de actualización	58265933216991557.

#### Duplicados y datos faltantes

En la base de datos correspondiente al mes de septiembre, se identificaron los siguientes casos de datos faltantes:

- Nivel Sisbén: 289.580 valores faltantes.
- Grupo poblacional del afiliado: 298.439 valores faltantes.

Además, al aplicar la función SHA256 para identificar registros duplicados, se detectaron 140 registros repetidos.

#### la unión de la base para el mes de octubre

Para el mes de octubre, se unieron las bases del sistema de salud y social correspondientes a los **regímenes contributivo y subsidiado**. Esta integración permite contar con la población total afiliada al sistema de seguridad social en salud en Colombia,

facilitando la identificación de tipos de datos, inconsistencias (como duplicados y valores nulos) y posibles estrategias de limpieza y manejo de la información.

VARIABLE	TIPO DE DATO	DESCRIPCION	
hash_sha256_num	Texto / cadena	Identificador único del afiliado	58265933216991557.
Género	Categórica	Sexo del afiliado	Femenino, Masculino
Grupo etario	Categórica / Rango	Edad en rangos por años	(1 a 5), (15 a 19), (50 a 55),ETC..
Código de la entidad	Texto / Código	Identificador de la EPS	EPS541,ETC..
Nombre de la entidad	Texto	Nombre completo de la EPS	NUEVA EPS S.A., EPS SURAMERICANA S.A EMSSANAR S.A.S.,ETC...
Régimen	Categórica	Tipo de régimen de afiliación	Contributivo, Subsidiado
Tipo de afiliado	Categórica	Rol del afiliado dentro del régimen	BENEFICIARIO, CABEZA DE FAMILIA
Estado del afiliado	Categórica	Estado actual de la afiliación	Activo, Protección Laboral C
Condición del beneficiario	Categórica	Condición especial del afiliado	NO APLICA, VICTIMAS DEL CONFLICTO ARMADO INTERNO, ETC...
Zona de Afiliación	Categórica	Tipo de zona donde reside	Urbana, Rural, Urbana-Cabecera Municipal
Departamento	Categórica	Departamento de residencia	CHOCÓ, ANTIOQUIA,ETC...
Municipio	Categórica	Municipio de residencia	ACANDI, SABANETA,ETC...
Nivel del Sisbén	Categórica / Entero	Nivel de puntaje del Sisbén	N,NAN,1,2,O,C,A,B,3,4
Grupo poblacional del afiliado	Categórica	Clasificación poblacional	POBLACION CON SISBEN, VÍCTIMAS DEL CONFLICTO ARMADO INTERNO, ETC...
Cantidad de registros	Numérica (entero)	Número de registros o unidades asociadas al afiliado (frecuencia / duplicados)	1, 2, 15, ETC...
Fecha de actualización	Fecha / Date Timé	Fecha y hora de la actualización del registro	2025-10-06 15:13:42
id fecha	Texto / Código	Identificador asociado a la	58265933216991557.

VARIABLE	TIPO DE DATO	DESCRIPCION	
		fecha de actualización	

### Duplicados y datos faltantes

En la base de datos correspondiente al mes de octubre, se identificaron los siguientes casos de datos faltantes:

- Nivel Sisbén: 17.044 valores faltantes.
- Grupo poblacional del afiliado: 35.150 valores faltantes.

Además, al aplicar la función SHA256 para detectar registros duplicados, se encontraron 280 registros repetidos.

### Estrategias de mitigación propuestas

Con base en los hallazgos identificados en las bases de datos correspondientes a los meses de septiembre y octubre, se proponen las siguientes estrategias de mitigación orientadas a mejorar la calidad, consistencia y completitud de la información registrada en la BDUA:

#### Datos faltantes

- Variable “Nivel Sisbén”
  - Implementar reglas de validación al momento del registro o carga de datos para evitar la omisión de este campo.
  - Complementar la información faltante mediante cruces con otras fuentes oficiales, como la base maestra del Sisbén IV administrada por el DNP.
  - En los análisis estadísticos, evaluar el impacto de los valores faltantes y aplicar técnicas de imputación cuando sea pertinente
- Variable “Grupo poblacional del afiliado”
  - Revisar los procesos de reporte de las EPS y entidades territoriales para garantizar el diligenciamiento completo de esta variable.
  - Implementar controles automáticos en el sistema de información que generen alertas cuando este campo se encuentre vacío.
  - En caso de análisis exploratorios, categorizar los valores faltantes como “No reportado” para mantener la trazabilidad.

#### Registros duplicados

- Aplicar filtros de deduplicación automatizada a través de identificadores únicos

- Validar los casos duplicados con las EPS responsables para confirmar si corresponden a registros legítimos
- Mantener únicamente el registro más actualizado según la **fecha de actualización** y conservar una bitácora de los cambios realizados para auditoría.