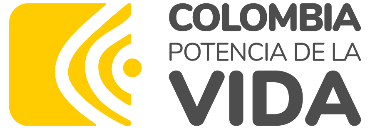


Informe de Migración, Validación y Cruce de Datos Catastrales: Terrenos, Construcciones y Unidades



Contenido

[1. Objetivos 5](#_Toc199374245)

[Objetivo general 5](#_Toc199374246)

[Objetivos específicos 5](#_Toc199374247)

[2. Metodología 6](#_Toc199374248)

[2.1 Criterios aplicados para el cruce de información 7](#_Toc199374249)

[2.1.1 Clasificación de Casos por Componente 7](#_Toc199374250)

[2.1.2 Secuencia Lógica de Tratamiento 8](#_Toc199374251)

[3. Migración de Información 8](#_Toc199374252)

[3.1. Terreno 8](#_Toc199374253)

[3.1.1 Migración 11](#_Toc199374254)

[3.1.2 Resultados 12](#_Toc199374255)

[3.1.3 Calidad 12](#_Toc199374256)

[3.2. Construcción 15](#_Toc199374257)

[3.2.1 Calidad 26](#_Toc199374258)

[3.2.2 Resultados 29](#_Toc199374259)

[3.3. Unidad 29](#_Toc199374260)

[3.3.1 Migración 35](#_Toc199374261)

[3.3.2 Calidad 39](#_Toc199374262)

[3.3.3 Resultados 44](#_Toc199374263)

Índice de tablas

[Tabla 1: Registros de terrenos Urbanos y rurales con sus informalidades 8](#_Toc199374285)

[Tabla 2: Depuración inicial de terrenos 8](#_Toc199374286)

[Tabla 3: Resultados de terrenos únicos y duplicados 9](#_Toc199374287)

[Tabla 4: Resumen de registros obtenidos para terreno 12](#_Toc199374288)

[Tabla 5: Caso 4 Resumen general de registros de Terreno entregados 42 municipios 12](#_Toc199374289)

[Tabla 6: Resultados de registros calidad 15](#_Toc199374290)

[Tabla 7: Tabla de comparaciones Caso 1 y Caso 2 en construcciones 18](#_Toc199374291)

[Tabla 8: Tabla de comparación Caso1. Caso 2 y Caso 3 en construcciones 19](#_Toc199374292)

[Tabla 9: Comparación de Caso 1, Caso 2, Caso 3 y Caso 4 en construcciones 20](#_Toc199374293)

[Tabla 10: Tabla de comparación de casos hasta el 5 en construcciones 22](#_Toc199374294)

[Tabla 11: Comparación de casos hasta el 6 en construcciones 23](#_Toc199374295)

[Tabla 12: Tabla de comparación del 1-7 en construcciones 24](#_Toc199374296)

[Tabla 13: Comparación de los casos 1-8 en construcciones 25](#_Toc199374297)

[Tabla 14: Coincidencia Emparejamiento aproximado solo por los primeros 22 dígitos del código, eligiendo el emparejamiento con menor diferencia de área 26](#_Toc199374298)

[Tabla 15: Resultados de casos del 1-8 en construcciones 26](#_Toc199374299)

[Tabla 16: Resultados del proceso de calidad de construcciones 29](#_Toc199374300)

[Tabla 17: Resumen de Unidades 29](#_Toc199374301)

[Tabla 18: Tabla Resumen casos de Unidades 35](#_Toc199374302)

[Tabla 19: Resúmen de casos del 1-12 39](#_Toc199374303)

[Tabla 20: Registros de unidades con código no válido 41](#_Toc199374304)

[Tabla 21: Resultados Finales de unidades 44](#_Toc199374305)

Índice de Ilustraciones

[Ilustración 1: Diagrama de Metodología 6](#_Toc199374306)

[Ilustración 2: Caso 1 Emparejamiento directo por número predial y código de Terrenos 9](#_Toc199374307)

[Ilustración 3: Caso 2 Agrupación terrenos que no tienen predio matriz y se les asigna el primer predio coincidente. 10](#_Toc199374308)

[Ilustración 4: Caso 3 Considera aquellos terrenos que no tienen relación con ningún predio. 11](#_Toc199374309)

[Ilustración 5: Caso 4 Recoge los registros descartados por duplicación de terrenos. 11](#_Toc199374310)

[Ilustración 6: Proceso de calidad - código predial corto 13](#_Toc199374311)

[Ilustración 7: Caso de duplicación 13](#_Toc199374312)

[Ilustración 8: Ejemplos de registro de Código de municipio no válido 14](#_Toc199374313)

[Ilustración 9: Ejemplos de Registros de Geometría no válida 14](#_Toc199374314)

[Ilustración 10: Ejemplos de Terreno sin predio asociado 15](#_Toc199374315)

[Ilustración 11: Ejemplos de Generación de construcciones por piso 16](#_Toc199374316)

[Ilustración 12: Empareja por código completo, piso, identificador y área ±5% en construcciones 17](#_Toc199374317)

[Ilustración 13: Igual a Caso 1, pero sin identificador, aplicando deduplicación por diferencia mínima de área en construcciones 18](#_Toc199374318)

[Ilustración 14: Coincidencia por Código, identificador y área ±5% en construcciones 19](#_Toc199374319)

[Ilustración 15: Coincidencia Código y área ±10% en construcciones 21](#_Toc199374320)

[Ilustración 16: Coincidencia en Código, piso e identificador (sin usar área) en construcciones 21](#_Toc199374321)

[Ilustración 17: Coincidencia de Código e identificador (sin piso ni área) 23](#_Toc199374322)

[Ilustración 18: Coincidencia Código y piso (sin identificador ni área) 24](#_Toc199374323)

[Ilustración 19: Registros con Código de municipio no válido 27](#_Toc199374324)

[Ilustración 20: Registros con Construcción sin identificador 28](#_Toc199374325)

[Ilustración 21: Registros con Construcciones sin piso o piso igual a cero 28](#_Toc199374326)

[Ilustración 22: Registros con Construcciones sin predio hasta la posición 22 29](#_Toc199374327)

[Ilustración 23:Caso 1 Emparejamiento exacto por código, identificador, piso y área 31](#_Toc199374328)

[Ilustración 24: Caso 2 Emparejamiento sin identificador por código, piso y área 31](#_Toc199374329)

[Ilustración 25: Caso 3 Emparejamiento por código e identificador con tolerancia de área (sin piso) 32](#_Toc199374330)

[Ilustración 26: Caso 4 Emparejamiento por código y área (sin piso ni identificador) 33](#_Toc199374331)

[Ilustración 27: Caso 5 Emparejamiento por código, identificador y piso (omitiendo área) 33](#_Toc199374332)

[Ilustración 28: Caso 6 Emparejamiento por código e identificador (omitiendo área y piso) 34](#_Toc199374333)

[Ilustración 29: Caso 7 Emparejamiento por código y piso (omitiendo identificador y área) 35](#_Toc199374334)

[Ilustración 30: Rescate por coincidencia de terreno y predial 36](#_Toc199374335)

[Ilustración 31: Comparación caso 8 y 9 36](#_Toc199374336)

[Ilustración 32: Unidades residuales sin geometría 37](#_Toc199374337)

[Ilustración 33: Migración de geometrías sin registro alfanumérico 38](#_Toc199374338)

[Ilustración 34: Registros de Unidades sin identificador 42](#_Toc199374339)

[Ilustración 35: Registro Unidades sin piso o piso cero 42](#_Toc199374340)

[Ilustración 36: Registros de Unidad sin predio 43](#_Toc199374341)

[Ilustración 37: Ejemplo de Unidades sin terreno asociado 44](#_Toc199374342)

# Objetivos

## Objetivo general

Ejecutar un proceso integral de migración, depuración, validación y emparejamiento de información geoespacial y alfanumérica de los componentes prediales —terreno, construcción y unidad—, con el fin de consolidar una base de datos limpia, estructurada y alineada con el modelo LADM\_COL, que sirva como insumo para la estructuración de capas oficiales en sistemas de información catastral.

## Objetivos específicos

* **Unificar** y estandarizar los datos geográficos y alfanuméricos de terrenos rurales y urbanos en una única estructura coherente y no duplicada.
* **Identificar y clasificar** los distintos escenarios de emparejamiento entre geometrías de terreno y predios (cruce directo, sin matriz, sin predio relacionado, duplicados), asignándoles una etiqueta de caso.
* I**ntegrar construcciones** y unidades con múltiples niveles de complejidad, aplicando reglas de cruce progresivas con base en código predial, área, identificador, número de piso y geometría.
* **Implementar validaciones de calidad** para detectar y documentar inconsistencias estructurales como códigos incorrectos, geometrías inválidas, objetos sin identificador o sin relación predial.
* **Consolidar la información migrada** en las tablas finales cr\_terreno y cr\_unidadconstruccion, con trazabilidad de origen y observaciones para análisis posterior.

# Metodología

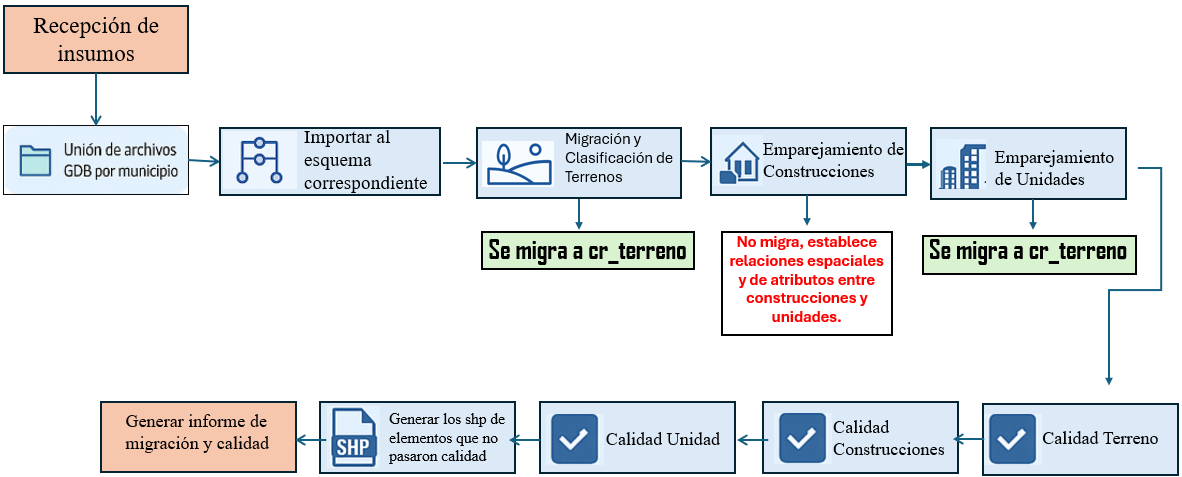


Ilustración : Diagrama de Metodología

El proceso inicia con la **consolidación de archivos geográficos (GDB) por municipio**, los cuales contienen la información espacial y alfanumérica necesaria para estructurar la base predial. Esta información se **importa a los esquemas correspondientes dentro de la base de datos**, según el tipo de componente que se esté procesando: terreno, construcción o unidad.

Una vez cargados los datos, se realiza la **normalización de la información de terrenos**, integrando tanto áreas urbanas como rurales en una estructura unificada. Como resultado de esta etapa, se genera una tabla depurada con registros únicos que luego son **migrados a la tabla final de terrenos**, correspondiente a la estructura oficial.

A continuación, se ejecuta el tratamiento de **las construcciones**, donde se desagregan por piso y se cruzan con las unidades funcionales asociadas. En esta fase no se realiza una migración directa de datos, ya que su propósito principal es **establecer relaciones geométricas y alfanuméricas entre las construcciones y las unidades**.

Seguido a esto, se lleva a cabo la **normalización de las unidades**, en la que se integran diferentes niveles de calidad y coincidencia con predios y construcciones. Los registros resultantes de este proceso son **migrados a la tabla final de unidades de construcción**, que almacena la información consolidada para uso oficial.

Después de la migración, se activa una **fase paralela de control de calidad**, que incluye la verificación de inconsistencias en códigos prediales, geometrías inválidas, duplicados, y la identificación de registros sin relación con predios. Esta revisión se realiza de forma independiente para terrenos, construcciones y unidades.

Con los resultados de calidad procesados, se generan los respectivos **archivos shapefile (SHP) a partir de ArcGIS**, organizados por capa temática, y posteriormente se **comprimen en formato ZIP** para su entrega o publicación.

## 2.1 Criterios aplicados para el cruce de información

La metodología aplicada para la migración y validación de la información predial se basó en una serie de **cruces progresivos y dependientes** entre los tres componentes fundamentales: **terreno, construcción y unidad**. El enfoque adoptado garantiza la trazabilidad, consistencia espacial y coherencia alfanumérica entre los registros, siguiendo una secuencia lógica que parte desde la base geográfica (terreno) hacia los componentes derivados.

Para establecer correspondencias válidas entre los datos geográficos y alfanuméricos de los distintos componentes, se aplicaron los siguientes criterios:

* **Código Predial Nacional (NPN):**  
  Se utilizó como criterio principal de enlace entre terrenos, predios y unidades. Los cruces directos se priorizaron por coincidencia exacta de los primeros 30 caracteres del código.
* **Identificador de la construcción o unidad:**  
  Permitió discriminar casos cuando existían múltiples elementos con el mismo código predial. Se utilizó especialmente en construcciones con varias unidades funcionales o pisos.
* **Número de Piso (Planta):**  
  Este campo fue estandarizado numéricamente para facilitar comparaciones y emparejamientos entre construcciones desagregadas por piso y unidades funcionales.
* **Área construida o del terreno (con tolerancia del ±5%):**  
  Se consideró una ventana de tolerancia en área para permitir cruces con leve variación métrica, producto de diferencias en trazado o digitación. Este criterio fue clave para la deduplicación y selección de candidatos más cercanos.
* **Geometría (Intersección o Punto Sobre Polígono):**  
  Se aplicaron funciones espaciales para verificar relaciones topológicas entre entidades, como ST\_Intersects o ST\_PointOnSurface, garantizando que construcciones y unidades se encuentren dentro del terreno que las contiene.

### 2.1.1 Clasificación de Casos por Componente

Cada componente fue depurado y migrado en función de una **clasificación de escenarios (casos)**, los cuales definen el nivel de coincidencia y la lógica aplicada para emparejar registros. Esta clasificación permite:

* Marcar casos exitosos de cruce directo.
* Identificar registros sin correspondencia o con ambigüedad.
* Aplicar reglas especiales para predios sin matriz o elementos duplicados.

Los casos están numerados (por ejemplo: Caso 1, Caso 2, etc.) y documentados en tablas separadas para trazabilidad y control.

### 2.1.2 Secuencia Lógica de Tratamiento

El flujo del proceso sigue una secuencia ordenada y dependiente entre componentes:

1. **Terreno**:  
   Constituye la base geométrica del sistema. Se normaliza, se eliminan duplicados y se migra según su coincidencia con predios.
2. **Construcción**:  
   Se descompone por número de pisos, se cruza con unidades funcionales y se prepara para servir como geometría base en casos donde las unidades no poseen geometría propia.
3. **Unidad**:  
   Se relaciona con terrenos y construcciones mediante código, identificador y geometría. Se clasifican en hasta 12 casos distintos según el nivel de coincidencia.

# Migración de Información

## 3.1. Terreno

* Fuentes de origen: datos urbanos y rurales

Tabla : Registros de terrenos Urbanos y rurales con sus informalidades

|  |  |  |
| --- | --- | --- |
| **Item** | **Urbano** | **Rural** |
| **Total Individual** | 189.420 | 219.256 |
| **Informal** | 1 | 1 |
| **Total** | 408.676 | |

El proceso de migración de terrenos definido en este script tiene como objetivo consolidar y clasificar la información de terrenos rurales y urbanos provenientes de las tablas **r\_terreno** y **u\_terreno**, alojadas en el esquema colsmart\_snc\_linea\_base. Inicialmente, se realiza un conteo de los registros, diferenciando entre terrenos formales e informales, y luego se combinan ambos conjuntos en una tabla unificada llamada z\_u\_r\_terreno. Esta tabla elimina duplicados basándose en una combinación del código predial y el área geométrica, conservando únicamente un registro representativo por terreno.

Tabla : Depuración inicial de terrenos

| **Bloque** | **Función principal** | **Propósito en el proceso** | **Criterios de depuración aplicados** |
| --- | --- | --- | --- |
| Depuración inicial | Consolidar terrenos únicos desde fuentes rurales y urbanas | Crear una tabla limpia y sin duplicados que sirva de base para la migración de terrenos | 1. Unifica las fuentes rurales y urbanas |
| 2. Calcula el área (sde.st\_area(shape)) |
| 3. Agrupa por codigo y área para identificar duplicados |
| 4. Conserva el registro con el t\_id (objectid) más alto como representativo |
| Registro de "caso 4" vinculado con la anterior tarea | Registrar los terrenos descartados por duplicación | Conservar trazabilidad de los registros no migrados por conflictos de duplicidad, para revisión o auditoría posterior | 1. Compara el conjunto original completo (terreno) con los registros finalmente seleccionados (z\_u\_r\_terreno) |
| 2. Identifica los t\_id que no fueron migrados (ausentes en la tabla de únicos) |
| 3. Registra estos duplicados con observación "Caso 4: Terreno no migrados por duplicados" |

Tabla : Resultados de terrenos únicos y duplicados

|  |  |
| --- | --- |
| **Terrenos únicos** | **Terrenos Duplicados** |
| 396.817 | 11.859 |

Posteriormente, los terrenos únicos se cruzan con la información de predios disponible en la tabla ilc\_predio, ubicada en el esquema colsmart\_test5\_owner, dando lugar a una clasificación en cuatro casos distintos:

#### Caso 1: Emparejamiento directo por número predial y código

Incluye los terrenos cuyo código coincide exactamente con un número predial nacional y puede migrarse directamente. Este caso representa los emparejamientos más directos y confiables entre terrenos y predios, basándose en una coincidencia exacta de código.

Los criterios aplicados son:

* Se seleccionan todos los registros de terreno cuya columna codigo coincide exactamente con el campo numero\_predial\_nacional de un predio.
* Esta coincidencia refleja una relación uno a uno clara y verificable, en la que el terreno y el predio comparten el mismo identificador predial completo.
* A estos registros se les asigna la observación "Caso 1", y se conservan datos clave como:
  + El id\_operacion del predio.
  + El globalid del predio.
  + La geometría (shape) y el t\_id del terreno.

Este caso constituye la base más sólida de emparejamiento dentro del proceso de migración o vinculación, y suele ser el punto de partida para los análisis posteriores.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso 1 Emparejamiento directo por número predial y código de Terrenos

#### Caso 2: Agrupa terrenos que no tienen predio matriz y se les asigna el primer predio coincidente.

Este caso identifica los terrenos cuyo código termina en '00000000', lo cual indica que no tienen una asociación directa con un predio específico. Además, se asegura que dichos terrenos no hayan sido ya clasificados en el Caso 1.

Para estos terrenos:

* Se busca el primer predio disponible (que no haya sido usado en Caso 1) que comparte los primeros 22 caracteres del número predial nacional con el código del terreno.
* Esta coincidencia parcial sugiere una relación jerárquica o agrupada dentro de una misma raíz predial.
* Al terreno se le asigna ese predio como una asociación aproximada, lo que permite mantener la trazabilidad de datos aunque no exista una coincidencia exacta.

Es importante tener en cuenta que, si hay más de un predio con la misma raíz, solo el primero alfabéticamente es asignado, y los demás no serán asociados a ese terreno en esta etapa.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso 2 Agrupación terrenos que no tienen predio matriz y se les asigna el primer predio coincidente.

#### Caso 3: Considera aquellos terrenos que no tienen relación con ningún predio.

Este script identifica los **terrenos que no han sido asociados a ningún predio**, es decir, aquellos que **no aparecen en los registros de los casos anteriores** (Caso 1 y Caso 2).

La lógica consiste en:

* Unir los identificadores (t\_id) ya registrados en terreno\_caso1 y terreno\_caso2.
* Seleccionar todos los terrenos de z\_u\_r\_terreno que **no están presentes en ninguno de esos casos previos**.
* Marcar dichos registros con la observación "Caso 3: Sin predio relacionado".

Este resultado permite detectar qué terrenos quedaron completamente fuera de las relaciones predio-terreno y podrían requerir una revisión especial o una asignación manual posterior.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso 3 Considera aquellos terrenos que no tienen relación con ningún predio.

#### Caso 4: Recoge los registros descartados por duplicación.

Como se indico al inicio de este apartado, este caso agrupa todos los registros de terreno que fueron descartados durante el proceso de consolidación por estar duplicados, ya sea por geometría, código o ambas condiciones. Representan los elementos que, aunque estaban en la fuente, no fueron seleccionados para la tabla final debido a que otro registro con la misma referencia fue priorizado.

**¿Cómo se identifican?**

1. Se parte de dos capas originales:
   * fuente Rural
   * fuente Urbana  
     **Cada registro recibe un t\_id único y se calcula su área (st\_area).**
2. Se construye una tabla *general, que contiene los terrenos únicos seleccionados para migración.*
3. Caso 4 identifica aquellos registros cuyo t\_id existe en las capas fuente pero no en la general ya depurada, usando una cláusula EXCEPT.  
   **Esto permite detectar todos los terrenos omitidos en la deduplicación.**
4. Finalmente, estos terrenos descartados se almacenan en la tabla z\_u\_r\_terreno\_caso4 con la observación: 'Caso 4: Terreno no migrado por duplicados'.

Tabla

El contenido generado por IA puede ser incorrecto.

De todos los terrenos con el mismo Código y area\_, selecciona **solo el que tiene el t\_id más alto**

Ilustración : Caso 4 Recoge los registros descartados por duplicación de terrenos.

### 3.1.1 Migración

**Migración de terrenos de Caso 1 y Caso 2**

* Se unifican los datos de caso 1 y caso 2.
* Se insertan en la tabla final cr\_terreno, incluyendo:
  + etiqueta, id\_operacion, predio\_guid, shape, codigo, identificador, observacion
* Se verifica la cantidad total de registros insertados y cuántos pertenecen específicamente al **Caso 2**.

**Migración de terrenos de Caso 3**

* Se insertan directamente desde caso 3 a cr\_terreno.
* En este caso no se migran **datos prediales** porque no existe relación (id\_operacion y predio\_guid no están presentes).
* Se generan igualmente los identificadores únicos (objectid, globalid) y se incluye la observación correspondiente.

### 3.1.2 Resultados

Resumen general de registros de Terreno entregados 42 municipios

Tabla : Resumen de registros obtenidos para terreno

|  |  |
| --- | --- |
| **Ámbito** | **Registros** |
| **Urbanos** | 189.420 |
| **Rurales** | 219.256 |
| **Total** | **408.676** |
| **Sin duplicar** | **396.817** |
| **Duplicados** | **11.859** |

Tabla : Caso 4 Resumen general de registros de Terreno entregados 42 municipios

|  |  |  |
| --- | --- | --- |
| **Informe de migracion** | | |
| **Caso** | **Cantidad** | **Descripción** |
| **1** | 391.351 | Migran por código (coincidencia exacta de identificador de terreno) |
| **2** | 2.254 | Terrenos marcados como **Predio Matriz sin relación con Predio** |
| **3** | 3.225 | Terrenos **sin predio relacionado** (falta de vínculo con tabla Predio) |
| **4** | 11.859 | Terrenos sin predio relacionado (falta de vínculo con tabla Predio) |
| **Total** | **408.689** | |

### 3.1.3 Calidad

Con el fin de garantizar la integridad y consistencia de la información de terrenos antes de su migración al modelo definitivo, se aplicaron múltiples controles de calidad sobre las capas fuente rurales y urbanas. Estas verificaciones fueron estructuradas en casos de observación, que permiten clasificar y documentar los registros con errores o condiciones especiales.

**a. Consolidación de capas fuente**

* Se unificaron las capas rurales y urbanas en una tabla intermedia denominada z\_b\_terreno.
* A cada terreno se le asignó un identificador único (t\_id) y se calculó el área de su geometría.
* Esta tabla fue la base para todos los análisis de calidad.

**b. Clasificación de casos de calidad**

Los registros observados fueron clasificados e insertados en la tabla z\_v\_terrenos con el campo caso, según el tipo de inconsistencia detectada:

* **Caso 1: Código predial corto**
* Se identificaron terrenos cuyo campo codigo tiene una longitud menor a 30 caracteres.
* Estos registros pueden corresponder a códigos incompletos o truncados.
* **Total detectado: 6 registros**

Imagen que contiene Interfaz de usuario gráfica

El contenido generado por IA puede ser incorrecto.

**NPN vacíos**

**29 dígitos**

Ilustración : Proceso de calidad - código predial corto

* **Caso 2: Terrenos duplicados**
* Se detectaron registros con el mismo codigo y la misma área geométrica.
* Se conserva solo el registro con menor t\_id, y los demás se marcan como duplicados.
* **Total detectado: 4.192 registros**

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso de duplicación

* **Caso 3: Código de municipio no válido**
* Se validó el código de municipio (primeros 5 caracteres del codigo) contra la tabla oficial igac\_municipios\_divipola.
* Se marcaron como inválidos aquellos códigos no presentes o erróneos.
* **Total detectado: 10 registros**

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Ejemplos de registro de Código de municipio no válido

* **Caso 4: Geometría no válida**
* Se marcaron los terrenos cuya geometría no corresponde a ST\_POLYGON o ST\_MULTIPOLYGON.
* Este tipo de geometría no es válida para superficies prediales.
* **Total detectado: 6 registros**

Texto

El contenido generado por IA puede ser incorrecto.

Ilustración : Ejemplos de Registros de Geometría no válida

* **Caso 5: Geometría nula**
* Terrenos sin geometría (shape IS NULL) fueron clasificados como inválidos.
* Estos registros deben ser corregidos o descartados.
* **Total detectado: 0 registros**
* **Caso 6: Terreno sin predio asociado**
* Se cruzaron los códigos de terreno con los predios (ilc\_predio) usando los primeros 22 caracteres.
* Aquellos terrenos que no encontraron coincidencia fueron clasificados como "sin predio".
* **Total detectado: 3.245**

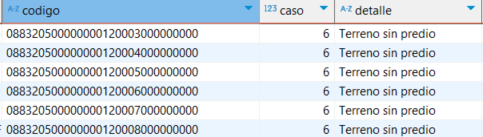


Ilustración : Ejemplos de Terreno sin predio asociado

#### 3.1.3.1 Resultados

Tabla : Resultados de registros calidad

|  |  |  |
| --- | --- | --- |
| **CASO** | **DETALLE** | **TOTAL** |
| 1 | longitud menor a 30 | 6 |
| 2 | Terreno duplicados | 11.853 |
| 3 | Cod de municipio no valido | 10 |
| 4 | Tipo de geometría no valido ST\_GEOMETRY | 6 |
| 6 | Terreno sin predio | 3245 |

## 3.2. Construcción

* **Generación de construcciones por piso**

Este procedimiento tiene como objetivo normalizar la información de construcciones rurales y urbanas para representar cada piso como una entidad independiente, y establecer su relación espacial con los terrenos disponibles.

1. **Entrada de datos**  
   Se parte de las tablas:
   * construcciones rurales
   * construcciones urbanas
2. **Normalización por número de pisos**  
   Cada construcción es descompuesta en una fila por piso
   * Si el número de pisos es nulo o cero, se asume que la construcción tiene un solo piso.
   * Si el número de pisos es mayor a 1, se generan n filas (una por cada piso).
3. **Cálculo de propiedades geométricas**  
   Para cada fila generada, se calcula:
   * El área
4. **Unión con terrenos**  
   Se construye un conjunto de terrenos rurales y urbanos (excluyendo aquellos con códigos informales donde el carácter 22 del código sea '2').  
   Se realiza una unión espacial entre las construcciones y los terrenos.
5. **Asignación del estado**
   * Si la construcción se logra emparejar con un terreno, se marca como 'base'.
   * Si no se encuentra un terreno correspondiente, se marca como 'nuevo'.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Ejemplos de Generación de construcciones por piso

* **Emparejamiento con unidades alfanuméricas**

Una vez generada la tabla anterior (una fila por piso con geometría y datos básicos), se realiza un proceso progresivo para emparejar estas construcciones con registros alfanuméricos de la tabla unidad\_data. Este proceso está estructurado por reglas jerárquicas llamadas **Casos 1 a 8**, donde cada caso aplica criterios distintos de comparación.

#### Caso 1: Empareja por código completo, piso, identificador y área ±5%.

En este primer paso se realiza la vinculación entre las construcciones geográficas (digitalizadas sobre el terreno) y la información alfanumérica registrada sobre unidades construidas. Para lograr un emparejamiento confiable, se aplican cuatro criterios simultáneamente:

1. **Código predial exacto:**  
   La construcción debe tener el mismo código predial que la unidad alfanumérica.
2. **Piso correspondiente:**  
   Se verifica que ambas estén ubicadas en el mismo nivel o piso. Esto se extrae de la información registrada en campo (por ejemplo, "Piso 1" o "Sótano 2") y se convierte en un valor numérico para comparar.
3. **Identificador único:**  
   Cada construcción tiene un identificador que debe coincidir con el de la unidad correspondiente. Este identificador asegura que se está comparando la misma entidad desde el punto de vista geográfico y administrativo.
4. **Área construida similar (±5%):**  
   Se compara el área medida en campo con el área geométrica calculada desde el polígono. Se aceptan pequeñas diferencias, siempre que la variación esté dentro del 5% hacia arriba o hacia abajo.

**Resultado obtenido:**  
Bajo estas condiciones estricts se logró emparejar **21.117** **construcciones** con sus respectivas unidades alfanuméricas. Estos registros representan los emparejamientos más confiables de todo el proceso, ya que cumplen con los cuatro criterios clave simultáneamente.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Empareja por código completo, piso, identificador y área ±5% en construcciones

#### Caso 2: Igual a Caso 1, pero sin identificador, aplicando deduplicación por diferencia mínima de área

En este segundo caso se amplía la cobertura de emparejamiento entre construcciones geográficas y unidades alfanuméricas, permitiendo casos en los que no existe identificador coincidente, pero se cumplen otras condiciones relevantes.

**Criterios aplicados:**

1. Código predial completo:  
   Se busca que el código predial de la construcción coincida exactamente con el de la unidad alfanumérica.
2. Piso correspondiente:  
   Se compara el piso de la construcción con el piso declarado en la unidad (limpiado para conservar solo el número).
3. Área similar (±5%):  
   Se acepta un margen de diferencia del 5% entre el área construida medida (alfanumérica) y el área geométrica digitalizada.

Tabla : Tabla de comparaciones Caso 1 y Caso 2 en construcciones

| **Criterio** | **Caso 1** | **Caso 2** |
| --- | --- | --- |
| **Código predial** | Coincide exactamente | Coincide exactamente |
| **Piso (nivel de la construcción)** | Coincide exactamente | Coincide exactamente |
| **Área construida** | Coincide con ±5% de diferencia | Coincide con ±5% de diferencia |
| **Identificador** | **Debe coincidir** | **No se requiere coincidencia** |
| **Deduplicación** | Se usa DISTINCT ON por código e identificador | Se escoge la coincidencia más cercana por área, evitando duplicados por objectid y id\_caracteristicasunidadconstru |
| **Precisión del emparejamiento** | Muy alta (exactitud total en todos los campos clave) | Alta, pero flexible (permite emparejamiento sin identificador) |
| **Cantidad de emparejamientos** | 21.117 | 1.352 |

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Igual a Caso 1, pero sin identificador, aplicando deduplicación por diferencia mínima de área en construcciones

#### Caso 3: Código, identificador y área ±5%

En este tercer caso se busca emparejar construcciones geográficas con unidades alfanuméricas cuando **no se dispone del dato confiable de piso**, pero sí coinciden el **código predial, el identificador y el área construida**.

Este caso permite ampliar la cobertura de emparejamientos cuando la información del nivel de la construcción (piso) está incompleta, mal registrada o no fue digitada.

**Criterios aplicados:**

1. **Código predial exacto:**  
   Se exige coincidencia total entre el código predial de la construcción y de la unidad alfanumérica.
2. **Identificador igual:**  
   El identificador de la construcción debe coincidir con el de la unidad registrada. Este campo sigue siendo clave para asegurar que se está hablando de la misma estructura.
3. **Área similar (±5%):**  
   Se acepta un margen de diferencia del 5% entre el área registrada en campo y el área geométrica calculada desde el polígono

**Lógica adicional**

* Este caso **no considera el piso** como condición de emparejamiento.
* Se aplica deduplicación:
  + Se escoge una única coincidencia por construcción (objectid).
  + Se escoge una única coincidencia por unidad (id\_caracteristicasunidadconstru).
  + Se prioriza la coincidencia cuya área es más cercana entre ambas fuentes.

**Resultado obtenido:**  
Se emparejaron correctamente **84.298 construcciones** que no pudieron ser cubiertas por los casos anteriores, pero que cumplen con las tres condiciones clave: código predial, identificador y área compatible

Tabla : Tabla de comparación Caso1. Caso 2 y Caso 3 en construcciones

|  |  |  |  |
| --- | --- | --- | --- |
| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — |
| **Identificador** | ✔️ | — | ✔️ |
| **Área construida ±5%** | ✔️ | ✔️ | ✔️ |
| **Deduplicación por área mínima** | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 |

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Coincidencia por Código, identificador y área ±5% en construcciones

#### Caso 4: Código y área ±10%

Este caso es el más flexible de los emparejamientos. Se aplica a las construcciones geográficas que **no pudieron ser vinculadas en los Casos 1, 2 o 3**, y busca emparejarlas con unidades alfanuméricas usando únicamente dos criterios:

**Criterios aplicados:**

1. **Código predial exacto:**  
   Se requiere coincidencia total entre el código predial de la construcción (codigo) y el de la unidad (numero\_predial).
2. **Área similar con margen ±10%:**  
   Se acepta que el área registrada en campo sea hasta un 10% mayor o menor que el área digitalizada.

**Características clave:**

* **No se considera el piso**.
* **No se exige coincidencia de identificador**.
* Es el emparejamiento más laxo, utilizado como último recurso.
* Se prioriza la coincidencia cuya diferencia de área sea la menor posible.
* Se evita duplicar emparejamientos ya usados en Casos anteriores.

**Resultado obtenido:**  
Este procedimiento permitió vincular **24.021 construcciones** adicionales que no fueron cubiertas anteriormente pero tienen una coincidencia razonable basada en código y área.

Tabla : Comparación de Caso 1, Caso 2, Caso 3 y Caso 4 en construcciones

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ | ✔️ |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — | — |
| **Identificador** | ✔️ | — | ✔️ | — |
| **Área construida ± tolerancia** | ±5% | ±5% | ±5% | **±10%** |
| **Deduplicación por área mínima** | ✔️ | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 | 24.021 |

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Coincidencia Código y área ±10% en construcciones

#### Caso 5: Código, piso e identificador (sin usar área)

Este caso se utiliza para intentar emparejar las construcciones restantes que **no fueron vinculadas en los casos anteriores (1 al 4)**. A diferencia de los anteriores, aquí **no se usa el área como criterio**.

**Criterios aplicados:**

1. **Código predial exacto:**  
   Se compara el código predial de la construcción con el de la unidad alfanumérica.
2. **Piso de la construcción:**  
   Se requiere coincidencia entre el número de piso registrado en la construcción y el piso indicado en la unidad (campo planta\_ubicacion).
3. **Identificador:**  
   También se exige que el identificador de la construcción coincida exactamente con el de la unidad alfanumérica.

**Particularidades:**

* **No se compara el área**, lo que lo diferencia de los casos anteriores.
* Se utiliza deduplicación por coincidencia más específica si hay más de una posible coincidencia.
* El piso se normaliza eliminando cualquier carácter no numérico para asegurar comparabilidad (Ej: "Piso 1A" → 1).

**Resultado obtenido:**  
Este caso permitió emparejar **7.328 construcciones** adicionales gracias a la coincidencia entre el código, el piso y el identificador, **sin necesidad de validar el área**.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Coincidencia en Código, piso e identificador (sin usar área) en construcciones

Tabla : Tabla de comparación de casos hasta el 5 en construcciones

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** | **Caso 5** |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — | — | ✔️ |
| **Identificador** | ✔️ | — | ✔️ | — | ✔️ |
| **Área construida comparada** | ✔️ ±5% | ✔️ ±5% | ✔️ ±5% | ✔️ ±10% | — |
| **Deduplicación por menor diferencia** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 | 24.021 | 7.328 |

#### Caso 6: Código e identificador (sin piso ni área)

En este caso se emparejan construcciones que no lograron ser vinculadas en los casos anteriores (del 1 al 5). Para ampliar la cobertura del emparejamiento, se aplica un criterio más flexible que prescinde del número de piso y del área construida como condición.

**Criterios utilizados:**

* Código predial exacto: Se comparan los códigos de la construcción y de la unidad alfanumérica.
* Identificador: Se requiere coincidencia exacta entre los identificadores de ambos elementos.
* No se utiliza el área como filtro de comparación, aunque se calcula la diferencia de área como criterio auxiliar para casos duplicados.

**Estrategia de deduplicación:**

Cuando varias unidades cumplen con los criterios anteriores, se conserva únicamente aquella que tenga la menor diferencia de área respecto a la construcción. Esto permite resolver ambigüedades sin imponer la validación por área como un requisito obligatorio.

**Resultado:**

Gracias a este método, se lograron emparejar 51.539 construcciones adicionales, incrementando significativamente la cobertura del proceso.

Tabla : Comparación de casos hasta el 6 en construcciones

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** | **Caso 5** | **Caso 6** |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — | — | ✔️ | — |
| **Identificador** | ✔️ | — | ✔️ | — | ✔️ | ✔️ |
| **Área construida comparada** | ✔️ ±5% | ✔️ ±5% | ✔️ ±5% | ✔️ ±10% | — | — |
| **Deduplicación por menor diferencia** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 | 24.021 | 7.328 | 51.539 |

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Coincidencia de Código e identificador (sin piso ni área)

#### Caso 7: Código y piso (sin identificador ni área)

Este caso busca emparejar construcciones y unidades alfanuméricas basándose únicamente en el código predial y el número de piso. No se exige coincidencia en el identificador ni en el área construida.

**Criterios aplicados:**

* Código predial exacto.
* Coincidencia del número de piso entre construcción y unidad (piso = planta\_ubicacion).
* Se excluyen los registros ya emparejados en los Casos 1 a 6.
* Se conserva solo la mejor coincidencia por construcción y unidad (deduplicación por menor diferencia de área, aunque esta no es criterio de emparejamiento).

**Resultado:**  
Este caso permitió emparejar **1.344 construcciones** adicionales que no habían sido asignadas previamente en los casos anteriores.

Tabla : Tabla de comparación del 1-7 en construcciones

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** | **Caso 5** | **Caso 6** | **caso 7** |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — | — | ✔️ | — | ✔️ |
| **Identificador** | ✔️ | — | ✔️ | — | ✔️ | ✔️ | — |
| **Área construida comparada** | ✔️ ±5% | ✔️ ±5% | ✔️ ±5% | ✔️ ±10% | — | — | — |
| **Deduplicación por menor diferencia** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 | 24.021 | 7.328 | 51.539 | 1.344 |

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Coincidencia Código y piso (sin identificador ni área)

#### Caso 8: Emparejamiento aproximado solo por los primeros 22 dígitos del código, eligiendo el emparejamiento con menor diferencia de área

Este criterio busca emparejar construcciones y unidades cuando no se ha logrado coincidencia en los casos anteriores. Para ello:

1. **Emparejamiento por código predial truncado**:  
   Se toman los primeros 22 caracteres del código predial de la construcción (codigo) y se comparan con los primeros 22 caracteres del número predial de la unidad (numero\_predial), permitiendo emparejamientos aproximados aunque el código completo no sea igual.
2. **Cálculo de diferencia de área**:  
   Para cada combinación posible entre construcción y unidad que coincide por los primeros 22 dígitos del código, se calcula la diferencia absoluta de área entre area\_ (de la construcción) y area\_construccion (de la unidad).
3. **Selección del mejor candidato**:  
   Se conservan solo los pares con menor diferencia de área:
   * Primero, se elimina la duplicación por objectid (seleccionando el menor diff\_area).
   * Luego, se asegura una única coincidencia por unidad (id\_caracteristicasunidadconstru), manteniendo la fila con menor diferencia.
4. **Exclusión de coincidencias previas**:  
   Este emparejamiento se realiza **únicamente** con aquellas construcciones y unidades que no fueron emparejadas en ninguno de los casos anteriores (del 1 al 7).

Tabla : Comparación de los casos 1-8 en construcciones

| **Criterio** | **Caso 1** | **Caso 2** | **Caso 3** | **Caso 4** | **Caso 5** | **Caso 6** | **caso 7** | **caso 8** |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Código predial exacto** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ Solo primeros 22 dígitos |
| **Piso (nivel de la construcción)** | ✔️ | ✔️ | — | — | ✔️ | — | ✔️ | — |
| **Identificador** | ✔️ | — | ✔️ | — | ✔️ | ✔️ | — | — |
| **Área construida comparada** | ✔️ ±5% | ✔️ ±5% | ✔️ ±5% | ✔️ ±10% | — | — | — | ✔️Solo para selección |
| **Deduplicación por menor diferencia** | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ | ✔️ |
| **Registros emparejados** | 21.117 | 1.352 | 84.298 | 24.021 | 7.328 | 51.539 | 1.344 | 15.457 |

Tabla : Coincidencia Emparejamiento aproximado solo por los primeros 22 dígitos del código, eligiendo el emparejamiento con menor diferencia de área

**Tabla

El contenido generado por IA puede ser incorrecto.**

Una vez ejecutados los 8 casos de emparejamiento entre construcciones y unidades prediales —cada uno aplicando criterios progresivamente más flexibles (código, piso, identificador, área, etc.)— se realiza una **consolidación final** de todos los registros emparejados.

Tabla : Resultados de casos del 1-8 en construcciones

|  |  |
| --- | --- |
| **Caso** | **Registros** |
| 1 | 21117 |
| 2 | 1352 |
| 3 | 84298 |
| 4 | 24021 |
| 5 | 7328 |
| 6 | 51539 |
| 7 | 1344 |
| 8 | 15457 |

### 3.2.1 Calidad

Con el fin de garantizar la integridad y consistencia de la información de construcciones antes de su migración al modelo definitivo, se aplicaron múltiples controles de calidad sobre las capas fuente rurales y urbanas. Estas verificaciones fueron estructuradas en casos de observación, que permiten clasificar y documentar los registros con errores o condiciones especiales que requieren atención o tratamiento diferenciado.

**a. Consolidación de capas fuente**

Se creó una tabla unificada, la cual consolida todas las construcciones provenientes de capas urbanas y rurales.Durante esta consolidación se asignó un nuevo identificador (t\_id) y se calculó el área geométrica de cada construcción mediante la función sde.st\_area(shape).

**b. Clasificación de registros según validaciones**

A continuación se describen los **casos de observación** aplicados a las construcciones:

* **Caso 1 – Código de municipio no válido**

Uno de los objetivos es detectar construcciones cuyo código predial contiene un código de municipio (los primeros 5 dígitos) que no coincide con la codificación oficial en la tabla igac\_municipios\_divipola.

Criterio aplicado:

* Se extraen los primeros 5 caracteres del campo codigo.
* Se validan contra los valores existentes en la columna mpcodigo de la tabla oficial de municipios.

Tratamiento:  
Los registros detectados se marcan con:

* caso = 1
* detalle = 'Código no válido'

**Cantidad de registros: 99**

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Registros con Código de municipio no válido

* **Caso 2 – Geometría de construcción no válida**

Este caso clasifica las construcciones cuya geometría no corresponde al tipo polígono (ST\_POLYGON). Estas geometrías atípicas pueden ser producto de errores en la digitalización o exportación de datos desde las capas originales. Se identifican aquellas construcciones con geometrías vacías, puntuales, lineales o de otro tipo diferente a polígono, y se excluyen del proceso de migración para evitar inconsistencias en la base de datos espacial. La clasificación incluye además una anotación con el tipo de geometría detectado para su posterior revisión.

Criterio aplicado:

* Se evalúa el tipo de geometría con la función sde.st\_geometrytype(shape) y se filtran todos los registros que no correspondan a 'ST\_POLYGON'.
* **Cantidad de registros: 2373**
* **Caso 3 – Construcción sin identificador**

Este caso agrupa las construcciones que no tienen un valor válido en el campo identificador, ya sea porque está vacío ('') o nulo (NULL). El identificador es un atributo clave que permite relacionar la construcción con su unidad correspondiente. La ausencia de este dato impide establecer vínculos alfanuméricos necesarios para la migración, por lo que estos registros deben ser separados para revisión y eventual corrección antes de su incorporación al modelo final.

**Criterio aplicado:**

* Se filtran los registros donde identificador IS NULL o identificador = ''.
* **Cantidad de registros: 1734**

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Registros con Construcción sin identificador

* **Caso 4: Construcciones sin piso o piso igual a cero**

Este caso identifica las construcciones que no tienen información registrada en el campo numero\_pisos o cuyo valor es igual a cero. Estas situaciones pueden indicar errores de captura o construcciones incompletas, lo que impide determinar la altura real o su distribución vertical. Esta información es fundamental para el emparejamiento correcto con unidades de construcción y para el modelado tridimensional posterior.

**Criterio aplicado:**  
Se seleccionan las construcciones que cumplen alguna de estas condiciones:

* **numero\_pisos IS NULL**
* **numero\_pisos = 0**
* **Cantidad de registros: 62065**

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Registros con Construcciones sin piso o piso igual a cero

* **Caso 5: Construcciones sin predio hasta la posición 22**

Se identifican construcciones cuyo código predial nacional (los primeros 22 caracteres del campo codigo) no encuentra coincidencia con ningún predio registrado en la tabla ilc\_predio. Estas construcciones no pueden asociarse a un predio válido y por lo tanto, se clasifican como registros con inconsistencia alfanumérica.

**Criterio aplicado**:

* Se comparó el inicio del código de la construcción (los primeros 22 dígitos) con los códigos de predio registrados.
* Si no se encontró ningún predio con ese mismo comienzo de código, la construcción se marcó como "sin predio asociado".
* Además, se descartaron los registros donde el código estaba vacío o era claramente inválido (por ejemplo, en blanco o igual a “0”).
* **Cantidad de registros: 1664**

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Registros con Construcciones sin predio hasta la posición 22

### 3.2.2 Resultados

Tabla : Resultados del proceso de calidad de construcciones

|  |  |  |
| --- | --- | --- |
| **Caso** | **Detalle** | **Registros** |
| 1 | Codigo no valido | 99 |
| 2 | Tipo de geometria no valido ST\_GEOMETRY | 2 |
| 2 | Tipo de geometria no valido ST\_MULTIPOLYGON | 2371 |
| 3 | Construccion sin identificador | 1734 |
| 4 | Sin pisos o piso cero | 62065 |
| 5 | Construcciones sin predio hasta la posicion 22 | 1664 |

## 3.3. Unidad

Tabla : Resumen de Unidades

|  |  |
| --- | --- |
| **Tipo de unidad** | **Cantidad** |
| Unidades rurales | 13794 |
| Unidades rurales nuevas | 852 |
| Unidades rurales informales | 512 |
| Unidades urbanas | 253588 |
| Unidades urbanas nuevas | 13952 |
| Unidades urbanas informales | 13087 |
| Total unidades | 267382 |

* **Generación de tabla única de unidades**

Este procedimiento tiene como objetivo **integrar y depurar la información de unidades rurales y urbanas**, creando una única tabla que contenga **solo los registros más representativos y sin duplicados**. Se realiza en varios pasos:

1. **Unión de datos rurales y urbanos**  
   Se combinan las unidades provenientes de zonas rurales y urbanas.  
   A las unidades urbanas se les suma un valor fijo al ID para diferenciarlas de las rurales.
2. **Cálculo del área y generación de ID temporal**  
   A cada unidad se le calcula su área geográfica y se le asigna un identificador interno (t\_id) que facilita el manejo de datos combinados.
3. **Selección de registros únicos**  
   Se agrupan los datos por **código predial, identificador, planta y área** para identificar registros duplicados.  
   De cada grupo se selecciona la versión más representativa usando el mayor t\_id.
4. **Normalización del campo de planta/piso**  
   Se limpia el valor de “planta” para extraer el número de piso. Si el valor está vacío o mal escrito, se le asigna el piso 1 por defecto.
5. **Creación de la tabla final**  
   Finalmente, se genera la tabla general, que contiene 259.261 unidades limpias, únicas y con atributos consistentes, listas para ser utilizadas en procesos de emparejamiento y análisis espacial.

Una vez consolidada la tabla única de unidades con geometría , el siguiente paso consiste en **relacionar estas geometrías con los registros alfanuméricos disponibles en unidad\_data**. Para ello, se establecen una serie de **reglas de emparejamiento** que permiten vincular ambos conjuntos de información.

* **Caso 1**

Representa el escenario ideal: se considera que una unidad alfanumérica tiene una coincidencia válida con una unidad geométrica cuando cumplen simultáneamente los siguientes criterios:

* El código predial (numero\_predial) coincide con el código de la geometría.
* El identificador de la unidad es el mismo.
* El piso o planta coincide después de limpiar el dato (por ejemplo, extrayendo el número de “Piso 2”).
* El área construida registrada está dentro de un rango aceptable respecto al área calculada de la geometría (±5% o ±1 m²).

Esta lógica permite generar una primera capa de emparejamientos confiables, que servirán como base para los casos siguientes, los cuales aplican reglas más flexibles o menos completas.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración :Caso 1 Emparejamiento exacto por código, identificador, piso y área

* **Caso 2**

Relacionar unidades alfanuméricas con geometría cuando no se dispone del identificador, pero se tiene coincidencia por:

* Código predial (numero\_predial = codigo)
* Piso (planta\_ubicacion ≈ planta)
* Área construida ≈ área geométrica (±5% o ±1 m²)

Importante:  
Se excluyen todas las coincidencias ya tomadas en el Caso 1.

Deducción de duplicados:

* Si hay más de una posible coincidencia para el mismo objectid, se queda solo con la de menor diferencia de área (diff\_area).
* Lo mismo aplica para el id\_caracteristicasunidadconstru.

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Caso 2 Emparejamiento sin identificador por código, piso y área

* **Caso 3**

Emparejar unidades alfanuméricas con sus geometrías cuando no se conoce con certeza el piso, pero sí hay coincidencia en:

* Código predial (numero\_predial = codigo)
* Identificador
* Y el área construida está dentro del rango permitido (±5% o ±1 m²)

Diferencia clave respecto a casos anteriores:

* Aquí ya no se exige que coincida el piso, a diferencia de los Casos 1 y 2.
* Se enfoca en coincidencias alfanuméricas por código e identificador, con tolerancia en área.

Filtrado:

* Se excluyen todas las coincidencias que ya fueron asignadas en los Casos 1 y 2.
* Luego, si hay más de una coincidencia posible, se conserva la que tiene menor diferencia de área (diff\_area), tanto por objectid como por id\_caracteristicasunidadconstru.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso 3 Emparejamiento por código e identificador con tolerancia de área (sin piso)

* **Caso 4**

Después de aplicar los tres primeros criterios de emparejamiento (Casos 1, 2 y 3), aún quedan registros que no han podido ser relacionados por falta de información en el campo de piso o identificador. En este contexto, el Caso 4 busca realizar coincidencias solo con base en el código predial y el área construida.

Para ello, se emparejan los registros alfanuméricos y geométricos que cumplan con lo siguiente:

* Coinciden en el código predial (numero\_predial = codigo).
* La área construida registrada está dentro de un rango aceptable respecto al área geométrica (±5% o ±1 m²).
* No han sido emparejados previamente en los Casos 1, 2 o 3.

Este caso no considera ni el piso ni el identificador debido a su ausencia o inconsistencias en los datos. Para evitar duplicaciones, en caso de múltiples coincidencias posibles se selecciona la de menor diferencia de área, tanto por geometría (objectid) como por unidad alfanumérica (id\_caracteristicasunidadconstru).

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Caso 4 Emparejamiento por código y área (sin piso ni identificador)

* **Caso 5**

El Caso 5 busca emparejar unidades alfanuméricas con geometrías cuando no fue posible validarlas por área, pero se cuenta con coincidencias en tres atributos clave:

* **Número predial (numero\_predial = codigo)**
* **Identificador**
* **Piso (planta\_ubicacion = planta luego de extraer el valor numérico)**

A diferencia de los casos anteriores, no se requiere que el área esté dentro de un rango aceptable. Esta estrategia se basa en el supuesto de que los datos alfanuméricos pueden ser lo suficientemente confiables en estos tres campos para establecer la relación.**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Caso 5 Emparejamiento por código, identificador y piso (omitiendo área)

* **Caso 6**

El Caso 6 busca relacionar unidades alfanuméricas con geometrías cuando ya no es posible utilizar la superficie construida (área) ni el piso como criterios de validación, pero aún se conserva una coincidencia directa en:

* **Código predial**
* **Identificador**

Esta estrategia se utiliza como último recurso lógico antes de recurrir a métodos exclusivamente espaciales. Puede deberse a errores de digitación en el piso, datos incompletos de área o ambigüedad entre varios registros con superficie similar.

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Caso 6 Emparejamiento por código e identificador (omitiendo área y piso)

* **Caso 7**

El Caso 7 representa una estrategia de cruce más relajada, pensada para aquellos registros alfanuméricos que no tienen identificador confiable y cuya área es inconsistente. Se recurre entonces a:

* **Código predial (numero\_predial = codigo)**
* **Piso (extraído y convertido numéricamente)**

Este enfoque busca capturar coincidencias válidas cuando el identificador está ausente, duplicado o mal digitado, pero el predio y el piso aún son consistentes. Aún se calcula la diferencia de área para seleccionar la mejor coincidencia posible, aunque no se filtra directamente por área.

**Tabla

El contenido generado por IA puede ser incorrecto.**

Ilustración : Caso 7 Emparejamiento por código y piso (omitiendo identificador y área)

* **Caso 8**

Este caso es el más laxo de todos los emparejamientos. Está diseñado para capturar coincidencias residuales cuando no se ha logrado emparejar una unidad con ninguna construcción mediante atributos más específicos. Se basa únicamente en:

* **Código predial (numero\_predial = codigo)**
* **No considera piso**
* **No considera identificador**

Aunque se calcula la diferencia de área para escoger la mejor coincidencia posible, el único criterio de cruce es el código predial. Sin embargo, es útil para salvar registros que de otro modo quedarían completamente sin emparejar.

### 3.3.1 Migración

Una vez implementados todos los casos de emparejamiento (del 1 al 8), se procede a consolidar los resultados en una única tabla. Este paso permite identificar cuántas unidades **no lograron ser migradas** mediante ninguno de los criterios establecidos.  
En este ejercicio, se detectaron **270.158 unidades sin emparejar**, las cuales requieren revisión adicional o serán tratadas por procesos alternativos.

Tabla : Tabla Resumen casos de Unidades

| **Caso** | **Código predial** | **Identificador** | **Piso** | **Área ±5%** | **Geometría** |
| --- | --- | --- | --- | --- | --- |
| 1 | ✅ | ✅ | ✅ | ✅ | ✅ |
| 2 | ✅ | ❌ | ✅ | ✅ | ✅ |
| 3 | ✅ | ✅ | ❌ | ✅ | ✅ |
| 4 | ✅ | ❌ | ❌ | ✅ | ✅ |
| 5 | ✅ | ✅ | ✅ | ❌ | ✅ |
| 6 | ✅ | ✅ | ❌ | ❌ | ✅ |
| 7 | ✅ | ❌ | ✅ | ❌ | ✅ |
| 8 | ✅ | ❌ | ❌ | ❌ | ✅ |

* **Caso 9**

Después de intentar emparejar todas las unidades con las construcciones en los casos anteriores (del 1 al 8), aún quedaron algunas sin relacionar. El caso 9 sirve para rescatar esas unidades.

Se revisa si esas unidades tienen un número predial válido y si ese número coincide con algún código de terreno en la base de datos. Si cumplen eso, se les asigna su geometría original (la forma que tenían antes) para que no se pierda su ubicación en el mapa.

Resultado: se recuperaron 95.683 unidades que no se habían migrado.

Imagen que contiene Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Rescate por coincidencia de terreno y predial

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Caso** | **Usa código predial** | **Compara área** | **Usa piso** | **Empareja con construcción** | **Geometría asignada** |
| 8 | ✅ Sí | ❌ No | ✅ Sí | ✅ Sí | De la construcción |
| 9 | ✅ Sí (5 dígitos) | ❌ No | ❌ No | ❌ No | Original de unidad |

Ilustración : Comparación caso 8 y 9

Después de identificado el Caso 9, que agrupa las unidades que no lograron emparejarse por geometría pero sí comparten el identificador técnico id\_caracteristicasunidadconstru, se procede a crear una tabla consolidada que reúne todos los registros migrados de los casos 1 al 9.

Esta tabla final representa el conjunto completo de unidades que lograron asociarse a una geometría de construcción, ya sea por coincidencia exacta, aproximada o mediante asignación directa por identificador. Su creación permite tener un panorama total de la migración y facilita la comparación con los registros originales para determinar el total de unidades aún pendientes por migrar.

* **Caso 10**

Una vez consolidada la información de los Casos 1 al 9 en la tabla, se procede con la clasificación de las unidades restantes que no lograron ser migradas con geometría, pero que aún cumplen ciertos criterios mínimos de pertenencia territorial.

Así, se crea la tabla del caso 10 la cual recopila las unidades completamente residuales. Estas no pudieron emparejarse ni siquiera por identificador (id\_caracteristicasunidadconstru) con una geometría de construcción, y por lo tanto se mantienen sin forma (shape = NULL), pero son marcadas con la observación "Caso 10" para fines de trazabilidad y revisión posterior.

Esta clasificación final es útil para:

* Saber exactamente qué unidades siguen pendientes de validación o ajuste.
* Permitir procesos posteriores de revisión manual o ajustes automáticos más flexibles.

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Unidades residuales sin geometría

* **Caso 11**

Una vez consolidados los casos del 1 al 10, se identifican aquellas **geometrías de unidades construidas** que **no lograron ser emparejadas con registros alfanuméricos**. Este caso garantiza que ninguna unidad geométrica válida quede excluida. Se migran con información mínima (solo atributos geométricos y de ubicación general).

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Migración de geometrías sin registro alfanumérico

* **Caso 12**

Este caso aborda construcciones nuevas que fueron identificadas en el cruce geométrico pero que **no se emparejaron con ninguna unidad del insumo original alfanumérico**. Se migran con los datos disponibles, priorizando mantener la geometría, y asignando valores por defecto a campos vacíos.

* **MIGRACIÓN A CR\_UNIDADCONSTRUCCION**

Antes de insertar las unidades migradas, es necesario limpiar los registros antiguos en la tabla destino cr\_unidadconstruccion. Esto se hace únicamente para los municipios involucrados en esta migración, identificados por los primeros 5 caracteres del campo código. Esta acción **evita duplicidad** y garantiza que la nueva información migrada no se mezcle con datos antiguos o incorrectos.

Luego de clasificar y consolidar los registros de unidades según los criterios de los casos 1 al 12, se procede a **migrar la información definitiva** al modelo de destino cr\_unidadconstruccion, respetando la naturaleza del cruce (alfanumérica, geométrica o combinada) y asegurando la integridad de las relaciones con predios y características.

**1. Migración de unidades con cruce completo (Casos 1 a 9)**

Estas unidades tienen información alfanumérica y geométrica completa, y cruzaron correctamente con predios y características. Se insertan con geometría (shape) y todos los atributos definidos.

**2. Migración de unidades solo alfanuméricas (Caso 10)**

Unidades que no tienen geometría, pero sí información alfanumérica y referencias válidas. Se migran sin campo shape.

**3. Migración de unidades solo con geometría (Caso 11)**

Estas unidades provienen de capas geográficas y no tienen coincidencia alfanumérica completa, pero sí geometría válida. Se generan identificadores artificiales y se migra con shape.

**4. Migración de construcciones nuevas no emparejadas (Caso 12)**

Registros con geometría de nuevas construcciones que **no se cruzaron con alfanumérico**. Se insertan directamente y se evita duplicidad con verificación por codigo y area.

**RESULTADOS**

Tabla : Resúmen de casos del 1-12

|  |  |
| --- | --- |
| **Caso** | **Registros** |
| Caso 1 | 20424 |
| Caso 2 | 629 |
| Caso 3 | 48262 |
| Caso 4 | 3923 |
| Caso 5 | 8233 |
| Caso 6 | 37718 |
| Caso 7 | 89 |
| Caso 8 | 2176 |
| Caso 9 | 95694 |
| Caso 10 | 140933 |
| Caso 11 | 127364 |
| Caso 12 | 138674 |

### 3.3.2 Calidad

Con el fin de garantizar la integridad y consistencia de la información de unidades antes de su migración al modelo definitivo, se aplicaron múltiples controles de calidad sobre las capas fuente rurales y urbanas. Estas verificaciones fueron estructuradas en casos de observación, que permiten clasificar y documentar los registros con errores o condiciones especiales que requieren atención o tratamiento diferenciado.

**a. Consolidación de capas fuente**

Se creó una tabla unificada, la cual consolida todas las construcciones provenientes de capas urbanas y rurales. Durante esta consolidación se asignó un nuevo identificador (t\_id) y se calculó el área geométrica de cada construcción mediante la función sde.st\_area(shape).

**b. Clasificación de registros según validaciones**

A continuación se describen los **casos de observación** aplicados a las construcciones:

* **Caso 1 – Duplicados por código, identificador, planta y área**

Este caso tiene como propósito detectar registros que se encuentran **duplicados** en la base de datos de unidades construidas, tomando como criterio de unicidad la combinación de los siguientes campos:

* código (número predial)
* identificador (letra asignada a la unidad)
* planta (nivel o piso de ubicación)
* área\_ (área geométrica calculada de la unidad)

Este tipo de duplicidad puede surgir por errores en la digitalización o por carga redundante de información durante el proceso de integración.

**Criterio aplicado:**  
• Se agrupan las unidades según los campos código, identificador, planta y área\_.  
• Se cuentan las ocurrencias por grupo.  
• Se marcan como duplicadas aquellas combinaciones con **más de una ocurrencia**.

**Tratamiento:**  
Los registros duplicados se insertan en la tabla z\_v\_unidad, conservando únicamente uno de ellos para los procesos posteriores (el de mayor t\_id).  
Los registros detectados se marcan con:  
• caso = 1  
• detalle = 'Unidad duplicada en código, piso, identificador, área'

**Cantidad de registros: 450**

* **Caso 2 – Código no válido**

Este caso identifica aquellas unidades cuyo **código predial** comienza con un código de municipio (los primeros 5 dígitos) que **no existe** en la tabla oficial igac\_municipios\_divipola. Es decir, detecta errores en el prefijo municipal del número predial.

Criterio aplicado:  
• Se extraen los **primeros 5 caracteres** del campo codigo.  
• Se comparan con los valores de la columna mpcodigo de igac\_municipios\_divipola (excepto el valor genérico '00000').  
• Se consideran “no válidos” aquellos códigos de municipio que no aparecen en la tabla oficial.

Tratamiento:  
Los registros detectados se insertan en la tabla de validación z\_v\_unidad con:

* caso = 2
* detalle = 'Codigo no valido'

**Cantidad de registros: 36**

Tabla : Registros de unidades con código no válido

Tabla

El contenido generado por IA puede ser incorrecto.

* **Caso 3 – Geometría no válida**

En este caso se detectan aquellas unidades cuya geometría no corresponde a un polígono (ST\_POLYGON), lo cual puede indicar errores en la captura o en la fuente de datos (por ejemplo, líneas, puntos o geometrías corruptas).

**Criterio aplicado:**

* Se toma cada geometría de la tabla unificada
* Se evalúa su tipo con sde.st\_geometrytype(shape).
* Se consideran inválidas todas las geometrías cuyo tipo **no** sea ST\_POLYGON (incluye ST\_POINT, ST\_LINESTRING, NULL, etc.).

**Tratamiento:**

* Los registros detectados se insertan en la tabla de validación z\_v\_unidad con:
  + caso = 3
  + detalle que concatena el texto "Tipo de geometria no valido " con el tipo real de la geometría (o "No definido" si es NULL).

**Caso 4 – Unidades sin identificador**

En este caso se detectan aquellas unidades que **no tienen asignado un identificador** (ya sea porque el campo está vacío o es NULL), lo cual impide distinguir correctamente unas unidades de otras cuando comparten el mismo predial, planta y área.

**Criterio aplicado:**

* Se toma cada registro de la tabla unificada
* Se filtran aquellos donde el campo identificador sea una cadena vacía ('') o esté NULL.

**Tratamiento:**

* Los registros detectados se insertan en la tabla de validación z\_v\_unidad con:
  + caso = 4
  + detalle = 'unidad sin identificador'

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Registros de Unidades sin identificador

**Caso 5 – Unidades sin piso o piso cero**

Este caso detecta aquellas unidades cuya información de piso (campo planta) está vacía, no contiene ningún dígito, o bien arroja 0 tras extraer los dígitos. Sin un piso válido, no se puede ubicar correctamente cada unidad en su nivel correspondiente.

**Criterio aplicado:**

1. Se limpia el texto de planta dejando sólo dígitos: REGEXP\_REPLACE(planta, '[^\d]+', '', 'g').
2. Se consideran inválidos los valores que quedan como cadena vacía ('') o que, al convertir a entero, resultan 0.

**Tratamiento:**  
Se insertan en la tabla de validación z\_v\_unidad con:

* caso = 5
* detalle = 'Unidad sin pisos o piso cero'

**Cantidad de registros detectados: 13.126**

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Registro Unidades sin piso o piso cero

**Caso 6 – Unidad sin predio**

En este caso se detectan aquellas unidades construidas cuyo código predial **no ha podido relacionarse** con ningún registro en la tabla de predios (ilc\_predio), usando los primeros 22 caracteres como criterio de emparejamiento.

**Criterio aplicado:**

* Se toman los 22 primeros caracteres del campo codigo de la unidad.
* Se comparan con los 22 primeros caracteres de numero\_predial\_nacional en la tabla ilc\_predio.
* Se consideran “sin predio” aquellas unidades para las que **no existe** ningún predio coincidente (p.id\_operacion IS NULL), excluyendo además códigos vacíos o triviales ('', ' ', '0').

**Tratamiento:**  
Los registros detectados se insertan en la tabla de validación z\_v\_unidad con:

* caso = 6
* detalle = 'Unidad sin predio hasta la posición 22'

**Cantidad de registros detectados: 3.966**

Tabla

El contenido generado por IA puede ser incorrecto.

Ilustración : Registros de Unidad sin predio

**Caso 7 – Unidades sin terreno asociado**

En este caso se detectan aquellas unidades que, al proyectar su punto central (st\_pointonsurface), **no intersectan** con ningún polígono de terreno (ni rural ni urbano). Esto indica que carecen de la referencia espacial de terreno a la que deberían pertenecer.

**Criterio aplicado:**

1. Se construye la tabla terreno unificando los terrenos rurales y urbanos, calculando además su área (area\_) y asignando un identificador único (t\_id).
2. Desde la tabla de unidades normalizadas, se hace un LEFT JOIN espacial con la tabla terreno, usando st\_intersects(st\_pointonsurface(unidad.shape), terreno.shape).
3. Se filtran las filas donde **no hay coincidencia** (t.t\_id IS NULL).

Gráfico

El contenido generado por IA puede ser incorrecto.

Terreno

unidad

Ilustración : Ejemplo de Unidades sin terreno asociado

### 3.3.3 Resultados

Tabla : Resultados Finales de unidades

|  |  |  |
| --- | --- | --- |
| **Caso** | **Descripción** | **Registros** |
| 1 | Unidad duplicada en codigo,piso,identificador,area | 2549 |
| 2 | Codigo no valido | 36 |
| 3 | Tipo de geometria no valido ST\_MULTIPOLYGON | 1555 |
| 4 | unidad sin identificador | 13126 |
| 5 | Unidad sin pisos o piso cero | 13126 |
| 6 | Unidad sin predio hasta la posicion 22 | 3966 |
| 7 | Unidad sin terreno asociado | 3478 |