

**DIRECCIÓN DE GEOESTADÍSTICA – DIG
GRUPO INTERNO DE TRABAJO INVESTIGACIÓN Y DESARROLLO-
GIT I+D**

**INFORME GENERAL
DEFINICIÓN DE AGRUPACIONES ESPACIALES
UNIDADES CENSALES CON ENFOQUE ÉTNICO**

DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA

OCTUBRE DE 2020



**El futuro
es de todos**

**Gobierno
de Colombia**

Contenido

SIGLAS	3
GLOSARIO	4
1. Introducción	5
2. Datos	6
3. Procesamiento	6
4. Resultados	8
1. Bibliografía	9

SIGLAS

CNPV 2018 Censo Nacional de Población y Vivienda del año 2018

DANE Departamento Administrativo Nacional de Estadística

DBSCAN *Density Based Spatial Clustering* (Clustering espacial basado en densidad)

DIG Dirección de Geoestadística

HDBSCAN *Hierarchical Density Based Spatial Clustering* (Clustering espacial basado en densidad jerárquica)

MGN Marco Geoestadístico Nacional

ML *Machine Learning* – Aprendizaje de Máquina

ONIC Organización Nacional Indígena de Colombia

VIHOPE Viviendas, Hogares y Personas

GLOSARIO

Density Based Spatial Clustering – DBSCAN (Agrupación espacial basada en densidad)

Técnica diseñada para descubrir las agrupaciones en una base de datos espacial, conociendo los parámetros apropiados y al menos un punto de cada *clúster* (Ester et al., 1996), proporcionando agrupaciones de objetos con base en un umbral de densidad global, y la similitud entre ellos, a partir de la distancia.

Sector rural¹

División cartográfica creada por el DANE para fines estadísticos. Es la mayor división definida para el área rural, tanto de los municipios como de los corregimientos departamentales. Corresponde en promedio a extensiones geográficas de 100 kilómetros cuadrados, cuyas áreas se encuentran delimitadas por límites político-administrativos, así como por elementos naturales y culturales del paisaje. Esta división contiene o agrupa secciones rurales.

¹ Manual de uso del Marco Geoestadístico Nacional en el proceso estadístico, volumen 2.0.

1. Introducción

Desde el año 2015, el Departamento Administrativo Nacional de Estadística –DANE, ha venido realizando procesos de Consulta Previa² con las organizaciones indígenas, en su momento para la realización del Censo Nacional de Población y Vivienda- CNPV 2018. Este operativo censal incluyó personal étnico, tanto censistas como supervisores, que garantizaron el enfoque diferencial para esta población, con el fin de evidenciar su realidad social y demográfica a través de la información estadística.

Actualmente, en el marco de la difusión y entrega de resultados del CNPV 2018, es necesario garantizar la anonimización y reserva estadística, específicamente, la georreferenciación de las viviendas donde al menos un integrante del hogar se auto reconoció como perteneciente a una comunidad étnica.

Consecuentemente, la Dirección de Geoestadística – DIG del DANE, desarrolló un proceso basado en técnicas de análisis de datos espaciales con el fin de generar agrupaciones (del inglés *clustering*) de las unidades censales con enfoque étnico, a partir de su distribución geográfica, obteniendo grupos de consistencia espacial, en función de la distancia y densidad de los puntos que representan estas unidades. Cada agrupación generada se representa por su centroide o punto medio en donde se agrupan los datos temáticos y estadísticos (viviendas, hogares y personas –VIHOPE) de las unidades que conforman esa agrupación. Cabe anotar que dicho centroide es un punto arbitrario que no coincide con la ubicación de las unidades censales, lo que garantiza la anonimización y reserva estadística antes mencionada. Adicionalmente, en los casos donde no se ejecutó el procedimiento técnico anterior, se optó por realizar las agrupaciones al sector rural, área definida en el Marco Geoestadístico Nacional del DANE.

En este documento se expone de manera resumida, los datos utilizados, la metodología propuesta y los resultados obtenidos a nivel nacional. Esta información busca ser de utilidad para diferentes actores, principalmente, la población étnica.

² <https://www.dane.gov.co/index.php/actualidad-dane/4464-el-dane-reitera-su-compromiso-con-las-organizaciones-indigenas>

2. Datos

A partir de los datos provenientes del CNPV 2018, se obtuvo información asociada a la unidad de observación (viviendas) mediante formularios, en donde el enfoque diferencial permitió identificar a la población perteneciente a alguno de los siguientes grupos étnicos: Indígenas; Gitanos o Rrom; Raizales del Archipiélago de San Andrés Providencia y Santa Catalina; Palenqueros de San Basilio; Negros, Mulatos, Afrodescendientes y Afrocolombianos.

Al agregar los datos temáticos y estadísticos de las unidades censales agrupadas espacialmente a sus correspondientes centroides, se busca garantizar la anonimización y reserva estadística que evite la individualización, no obstante, facilitando la difusión y disposición de resultados a un adecuado nivel de desagregación.

3. Procesamiento

A continuación, se describe, en términos generales, el proceso realizado:

1. Acopio o recepción de las unidades censales del CNPV 2018 con información asociada de autorreconocimiento étnico.
2. Implementación del algoritmo de *Machine Learning* HDBSCAN para la creación de agrupaciones espaciales para cada municipio, integradas por mínimo tres (3) unidades (criterio temático establecido para evitar la identificación de unidades puntuales), utilizando los datos seleccionados previamente.
3. Revisión de calidad de resultados. Para ello se inspecciona la conformación y consistencia espacial de las agrupaciones generadas, verificando su ubicación geográfica dentro de la entidad territorial, junto a la validación del cumplimiento del criterio temático de agrupación.
4. En los municipios donde no se aplicó el algoritmo de agrupación espacial por situaciones asociadas a la cantidad y distribución de unidades censales, éstas serán asociadas y agrupadas al centroide del sector rural del MGN, donde se ubiquen.
5. Generación de cobertura geográfica de centroides de las agrupaciones espaciales de unidades censales con enfoque étnico- 2018. Esta cobertura se remite a la Dirección de Censos y Demografía para la correspondiente asociación de datos temáticos y estadísticos.

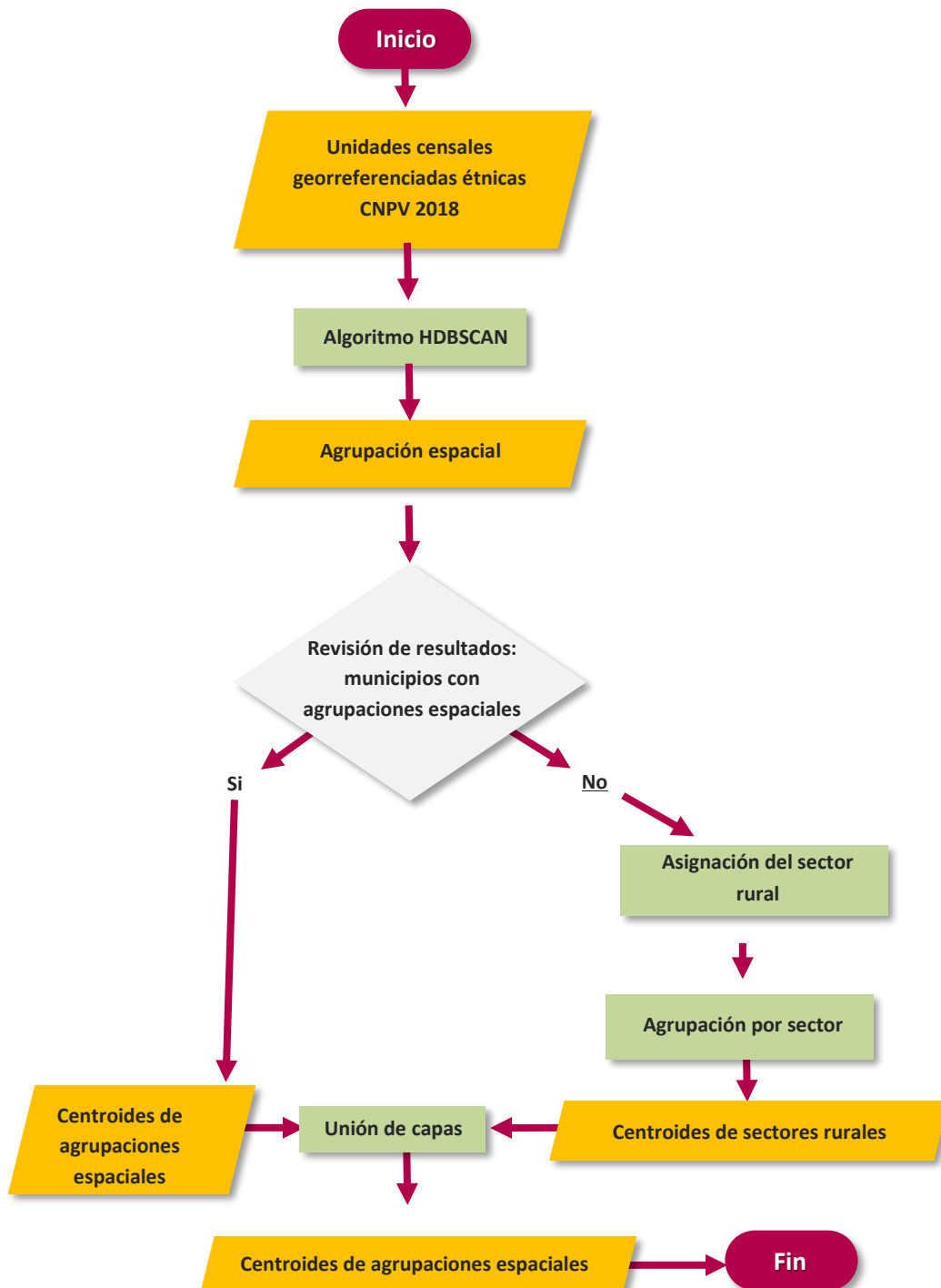


Ilustración 1 Flujo del proceso realizado para la generación de agrupaciones espaciales

4. Resultados

Las siguientes ilustraciones representan la forma de implementación de la técnica de agrupación espacial empleada (HDBSCAN).

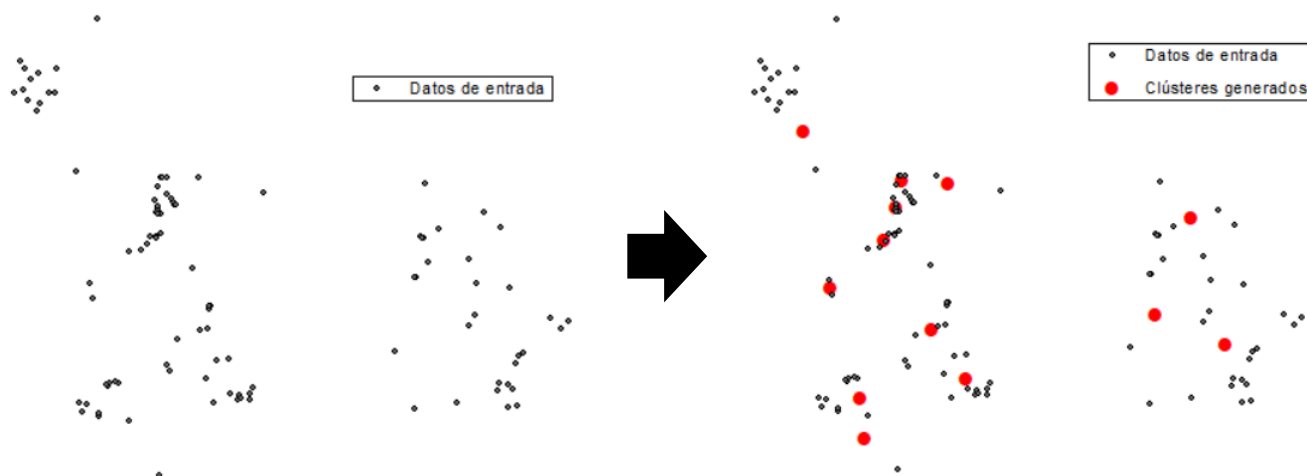


Ilustración 2 Ejemplo de agrupación de datos CNPV 2018

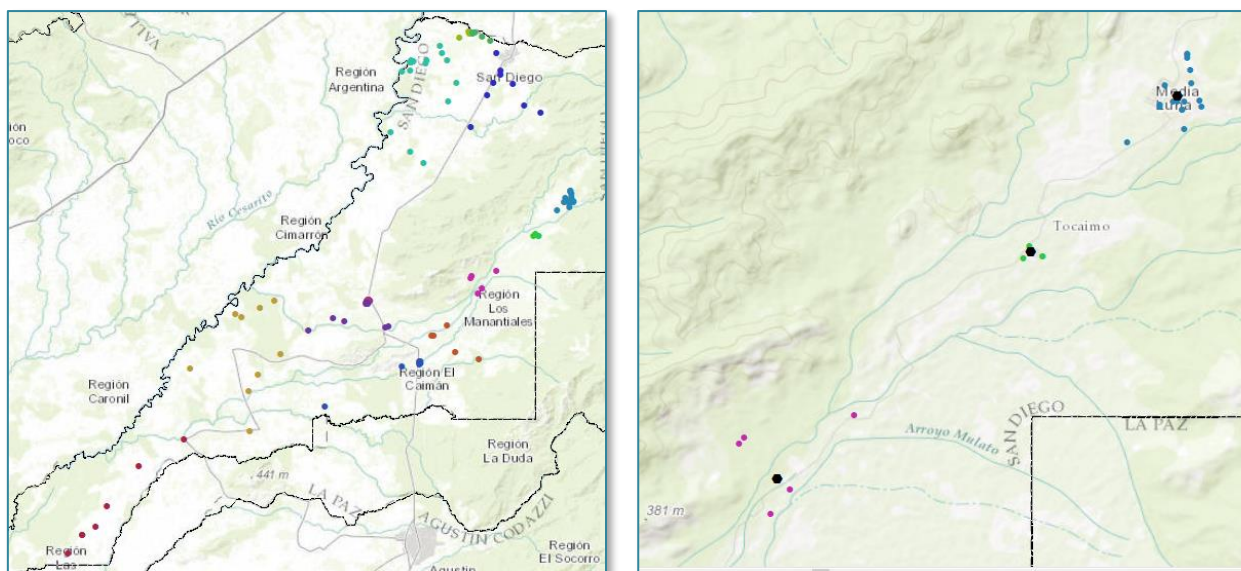


Ilustración 3 Representación de agrupaciones y ampliación que visualiza los centroides por cada agrupación

Bibliografía

- DANE. (2020). *COLOMBIA - Censo Nacional de Población y Vivienda—CNPV - 2018*. Departamento Administrativo Nacional de Estadística. <http://microdatos.dane.gov.co/index.php>
- Ester, M., Kriegel, H.-P., & Xu, X. (1996). *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. 6.
- Kaufman, L., & Rousseeuw, P. (1990). *Finding Groups in Data An Introduction to Cluster Analysis*.
- McInnes, L., Healy, J., & Astels, S. (2017). HDBSCAN: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205. <https://doi.org/10.21105/joss.00205>