

# Segmentación de Municipios en Colombia según Acceso a Servicios Públicos y Variables de Salud

## 1. Resumen

En este proyecto, se analizarán los datos de acceso a servicios públicos y variables de salud a nivel municipal en Colombia utilizando técnicas de aprendizaje no supervisado. Se emplearán métodos como el clustering jerárquico y K-means para identificar grupos de municipios con características similares en términos de acceso a servicios y salud pública. El objetivo es proporcionar insights valiosos que puedan ayudar a los responsables de políticas públicas, como el Ministerio de Salud y Protección Social, y los departamentos regionales, a focalizar recursos y mejorar la planificación estratégica en diferentes regiones del país. A través del análisis de clusters, se espera descubrir patrones ocultos y segmentar los municipios en grupos homogéneos, permitiendo una mejor toma de decisiones en la asignación de recursos y el diseño de intervenciones. Este proyecto contribuirá al campo al proporcionar una base de datos segmentada que facilitará la identificación de áreas con necesidades críticas y potencial para intervenciones dirigidas.

## 2. Introducción

El acceso equitativo a servicios públicos y de salud es fundamental para el desarrollo sostenible y el bienestar de las comunidades. En Colombia, la disparidad en el acceso a estos servicios entre los municipios es considerable, lo que representa un desafío significativo para los responsables de políticas públicas. Según datos del Departamento Administrativo Nacional de Estadística (DANE), las variaciones en la cobertura de servicios públicos son marcadas, con algunos municipios donde la cobertura apenas supera el 50%, mientras que en otros llega hasta el 90%, lo que evidencia profundas desigualdades socioeconómicas y geográficas([DANE](#))

Este proyecto utiliza técnicas de aprendizaje no supervisado como el clustering jerárquico, K-means, DBSCAN y GMM para descubrir patrones ocultos y segmentar los municipios de acuerdo con su acceso a servicios y variables de salud. Estas herramientas permiten identificar grupos de municipios con necesidades similares, facilitando la toma de decisiones informadas para la asignación efectiva de recursos. A través de este enfoque, se pretende ofrecer recomendaciones basadas en datos que mejoren la planificación y ejecución de políticas públicas, particularmente en contextos donde los recursos son limitados y es crucial priorizar las intervenciones.

Aplicando estos métodos avanzados de análisis de datos a un conjunto exhaustivo de indicadores municipales, hemos identificado varios clusters con características distintas, lo que proporciona una valiosa perspectiva para la focalización de intervenciones. A pesar de los desafíos relacionados con la calidad y completitud de los datos, los resultados son prometedores y sugieren que una estrategia dirigida puede ser más efectiva que la aplicación de políticas generales. Con base en estos hallazgos, recomendamos que las entidades gubernamentales adopten un enfoque de monitoreo y evaluación continua, utilizando las técnicas demostradas aquí para adaptar las políticas públicas a las necesidades específicas y cambiantes de los municipios colombianos.

### Objetivos del Proyecto:

- Identificar grupos de municipios con características similares en términos de acceso a servicios públicos y salud.
- Proporcionar recomendaciones basadas en los resultados para mejorar la asignación de recursos y la implementación de políticas públicas.
- Aplicar técnicas de aprendizaje no supervisado para el análisis de datos socioeconómicos y de salud.

### 3. Revisión Preliminar de la Literatura

En la última década, el uso de técnicas de aprendizaje no supervisado ha crecido significativamente en el análisis de datos socioeconómicos y de salud. Estudios previos han demostrado la utilidad del clustering para identificar patrones en datos geográficos y demográficos. Por ejemplo, un informe de la Organización para la Cooperación y el Desarrollo Económicos (OCDE) utilizó análisis de clusters para evaluar la transferibilidad de intervenciones en salud pública, lo que permitió identificar áreas con necesidades críticas . Otro estudio relevante de la National Institutes of Health (NIH) también explora la utilización de estas técnicas en el ámbito de la salud pública, destacando cómo la combinación de clustering y PCA ha sido útil para reducir la complejidad en los análisis de bienestar social y facilitar la visualización de los datos . Estos estudios proporcionan una base sólida para la metodología aplicada en este proyecto, que busca explorar patrones y segmentar municipios colombianos con base en indicadores de acceso a servicios públicos y de salud.

### 4. Materiales y Métodos

Este proyecto se llevó a cabo utilizando datos proporcionados por el Departamento Administrativo Nacional de Estadística (DANE) y otras fuentes relevantes. El conjunto de datos incluía información a nivel municipal sobre cobertura de servicios públicos (como agua potable, alcantarillado, energía eléctrica) y variables relacionadas con la salud (como mortalidad infantil, cobertura de regímenes de salud, y nacidos vivos).

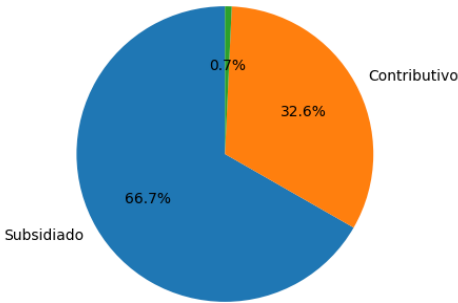
#### Fuentes de Datos:

- **DANE:** Proporcionó los datos sobre la cobertura de servicios públicos y variables demográficas para los municipios en Colombia.
- **Panel de Salud y Servicios Públicos a Nivel Municipal:** Datos adicionales que incluían variables relacionadas con salud pública en los municipios colombianos.

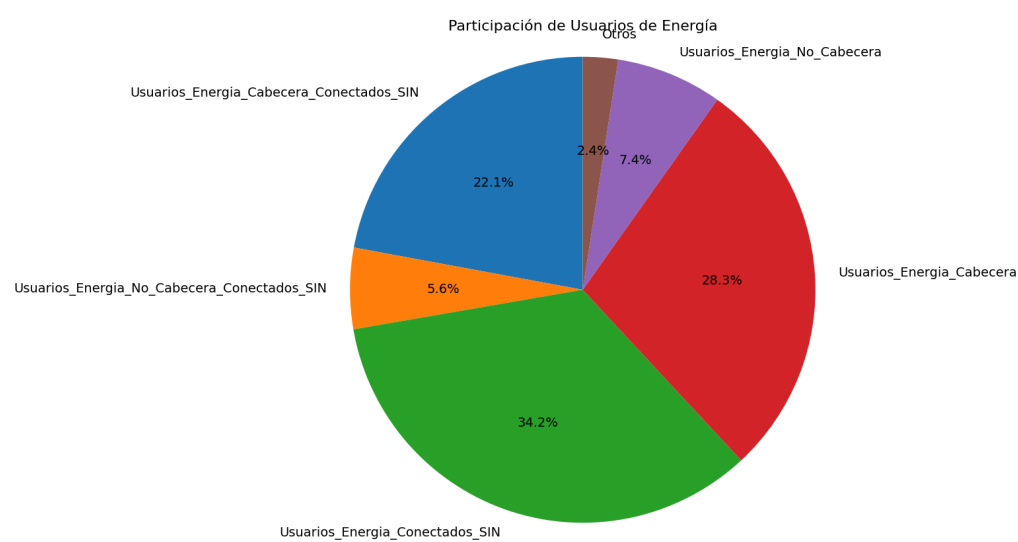
#### Análisis descriptivo

	Cobertura_Regimen_Subsidiado	Cobertura_Acueducto_Total	Cobertura_Acueducto_Urbana	Cobertura_Acueducto_Rural	Cobertura_Aseo_Total	Cobertura_Aseo_Urbana	
count	20936.000000	10391.000000	10391.000000	10391.000000	9293.000000	9293.000000	count
mean	0.889407	59.040557	83.718009	36.435894	45.008191	79.860396	mean
std	0.282067	30.349506	25.140655	37.048641	30.522410	27.975329	std
min	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	min
25%	0.730430	33.380000	74.655000	0.040000	19.580000	68.020000	25%
50%	0.963107	59.003690	98.010000	24.450000	38.350000	94.360000	50%
75%	1.000000	87.857910	100.000000	68.310000	67.740000	100.000000	75%
max	3.129962	100.000000	100.000000	100.000000	100.000000	100.000000	max
	Cobertura_Alcantarillado_Total	Cobertura_Alcantarillado_Urbana	Cobertura_Alcantarillado_Rural	Cobertura_Potencial_Gas_Natural	Cobertura_Efectiva_Gas_Natural		
	10391.000000	10391.000000	10391.000000	5206.000000	5206.000000		
	40.602562	75.013455	12.891793	89.254990	70.502007		
	30.065353	32.140062	25.586621	45.973456	52.378305		
	0.000000	0.000000	0.000000	0.000000	0.000000		
	15.768377	61.115000	0.000000	82.529690	54.146544		
	33.980000	90.883191	0.330000	94.770813	74.160255		
	62.069483	100.000000	11.615000	99.469025	87.273790		
	100.000000	100.000000	100.000000	2685.935547	3009.715820		

En la cobertura que emb



El régimen con mayor participación es el subsidiado, al cual pertenece aproximadamente dos tercios de la población. Por otro lado, el régimen especial representa menos del 1% de participación.



El gráfico muestra una clara división entre usuarios conectados al **Sistema Interconectado Nacional (SIN)**, que representa una gran mayoría, y los que están fuera de la cabecera municipal y no conectados a este sistema. La cabecera municipal (áreas urbanas) tiene una participación importante, tanto en términos de usuarios conectados al SIN como no conectados, mientras que los usuarios fuera de la cabecera tienen una menor representación.

Algoritmos Utilizados:

- **Clustering Jerárquico:** Se implementó el clustering jerárquico utilizando el método de enlace de Ward para minimizar la varianza dentro de los grupos. Se utilizó un dendrograma para visualizar la estructura jerárquica.
- **K-means:** Se utilizó K-means con 4 clusters. El número de clusters fue determinado utilizando el índice de silueta y el método del codo.
- **DBSCAN:** DBSCAN fue útil para identificar municipios que presentaban comportamientos atípicos o outliers.
- **GMM (Gaussian Mixture Model):** Se implementó GMM para capturar la incertidumbre en las asignaciones de los clusters, lo que proporcionó una visión probabilística de la pertenencia a clusters.

5. Resultados y Discusión

La aplicación de diferentes algoritmos de clustering nos permitió identificar grupos de municipios con características comunes en cuanto a su acceso a servicios públicos y variables de salud. Cada uno de los algoritmos utilizados brindó una perspectiva diferente sobre cómo los municipios se agrupan, destacando patrones relevantes para la planificación de políticas públicas.

Resultados Clave

- **Clustering Jerárquico:** El dendrograma generado mediante el método de enlace de Ward mostró la formación de tres clústeres principales. Cada uno de estos clústeres representa un nivel distinto de acceso a los servicios públicos y de salud. Este enfoque permitió identificar no solo los grupos generales, sino también subgrupos dentro de los clústeres principales, lo que ofrece una visión más matizada de la estructura interna de los datos. Aunque no se especificaron automáticamente el número de clústeres, fue útil para explorar la jerarquía de similitudes entre los municipios.
- **K-means:** Utilizando un enfoque de K-means con cuatro clústeres, identificamos que los municipios con menor acceso a servicios básicos forman un grupo homogéneo y bien definido. El análisis mediante PCA (Análisis de Componentes Principales) mostró que la

mayoría de los municipios se agrupan en torno a tres clústeres principales, con un cuarto clúster que representa outliers con características extremas. Los municipios del **Clúster 3** presentan acceso significativamente limitado a servicios, mientras que los del **Clúster 2** se destacaron por tener acceso desproporcionadamente alto en comparación con el resto.

- **DBSCAN**: Este algoritmo fue particularmente útil para identificar outliers en los datos. El clúster principal (**Clúster 0**) incluye la mayoría de los municipios con características más comunes, mientras que el **Clúster -1** identificó 77 municipios como outliers, que presentan características muy atípicas en términos de acceso a servicios públicos. Estos outliers representan municipios con acceso extremadamente bajo o anómalo a ciertos servicios, lo que podría sugerir zonas con problemáticas especiales o intervenidas por políticas específicas.
- **Gaussian Mixture Model (GMM)**: GMM identificó cuatro clústeres principales, similares a los obtenidos por K-means, pero con la ventaja adicional de asignar probabilidades a cada municipio sobre su pertenencia a diferentes clústeres. Esto permite una interpretación más flexible, ya que los municipios no son asignados de manera determinista a un solo clúster. En particular, el **Clúster 3**, con 4,799 municipios, representó a los municipios con mayores dificultades, mientras que el **Clúster 2** nuevamente identificó los municipios con un acceso desproporcionado a servicios.

## Discusión de Resultados

- **Comparación de algoritmos**: Tanto K-means como GMM produjeron resultados consistentes, identificando cuatro clústeres bien diferenciados. Estos clústeres agrupan a municipios según su acceso a servicios de salud, agua potable, electricidad y alcantarillado, entre otros. Mientras que K-means ofrece una partición más clara y determinista, GMM permite una interpretación más flexible, al asignar probabilidades a los municipios que pertenecen a múltiples clústeres.
- **Clustering Jerárquico**: Aunque el clustering jerárquico no especificó automáticamente un número de clústeres, fue útil para explorar cómo los municipios se agrupan jerárquicamente, ofreciendo una visión más detallada de las relaciones entre ellos. Este enfoque puede ayudar a identificar subgrupos más específicos dentro de los clústeres principales, lo cual es útil para intervenciones más dirigidas.
- **DBSCAN y la identificación de outliers**: DBSCAN sobresalió al identificar municipios que se comportan como outliers. Estos municipios, especialmente aquellos agrupados en el **Clúster -1**, representan casos atípicos, lo que sugiere que requieren una mayor atención por parte de los responsables de políticas públicas. Estos municipios pueden estar enfrentando dificultades críticas en su acceso a los servicios o pueden estar sujetos a intervenciones específicas que los distinguen de los municipios restantes.

## 6. Conclusión

**Conclusiones Generales del Análisis de Clustering**: El análisis de clustering reveló patrones significativos en la distribución de servicios públicos y variables de salud entre los municipios de Colombia. Se emplearon diferentes métodos de clustering: K-means, DBSCAN, y Gaussian Mixture Model (GMM), cada uno ofreciendo perspectivas únicas sobre la agrupación de los municipios basada en múltiples variables socioeconómicas y de infraestructura.

1. **K-means** mostró una distribución relativamente uniforme entre dos clusters principales, lo que sugiere una división clara entre los municipios con servicios bien establecidos y aquellos en desventaja.
2. **DBSCAN**, por otro lado, identificó un grupo predominante y varios grupos menores, reflejando una alta densidad de municipios con características similares y algunos outliers que podrían tener condiciones únicas.
3. **GMM** presentó una clara separación en cuatro grupos, proporcionando una subdivisión más matizada que puede ser crítica para la implementación de políticas específicas y dirigidas.

**Recomendaciones Basadas en los Resultados:**

- **Cluster 0 ("Altamente Servido"):** Incluye municipios con excelente acceso a servicios y altos estándares de salud. Recomendamos mantener y mejorar las políticas actuales en estos lugares.
- **Cluster 1 ("Medianamente Servido"):** Municipios con servicios adecuados pero mejorable. Se sugiere focalizar intervenciones que fortalezcan la infraestructura existente.
- **Cluster 2 ("Servicio Bajo"):** Municipios que requieren atención urgente para mejorar el acceso a servicios básicos y atención médica.
- **Cluster 3 ("Críticamente Desatendido"):** Representa municipios con severas deficiencias en todos los ámbitos evaluados. Es crucial una intervención estatal inmediata para evitar una crisis humanitaria.

**Implicaciones para Políticas Públicas:** La identificación de estos clusters permite a las entidades gubernamentales dirigir recursos de manera más efectiva, priorizando áreas con necesidades urgentes y optimizando las intervenciones para maximizar su impacto. Los datos sugieren que las políticas no deben ser uniformes, sino adaptadas a las características específicas de cada cluster de municipios.

**Futuras Direcciones de Investigación:** Es esencial continuar este análisis a medida que se disponga de nuevos datos, utilizando los clusters identificados como base para un monitoreo dinámico y adaptativo de las condiciones y necesidades municipales. También se recomienda explorar modelos predictivos que puedan anticipar cambios en los clusters basados en tendencias socioeconómicas y climáticas.

En conclusión, el estudio proporciona una base sólida para una toma de decisiones informada y basada en datos, esencial para el desarrollo equitativo y sostenible de Colombia.

**7. Bibliografía**

1. Departamento Administrativo Nacional de Estadística (DANE). (2023). "Pobreza monetaria y desigualdad en Colombia." Recuperado de <https://www.dane.gov.co/index.php/estadisticas-por-tema/pobreza-y-condiciones-de-vida/pobreza-y-desigualdad>.
2. Departamento Administrativo Nacional de Estadística (DANE). (2021). "Comunicado de Pobreza Monetaria 2021." Recuperado de [https://www.dane.gov.co/files/investigaciones/condiciones\\_vida/pobreza/2021/Comunicado-pobreza-monetaria\\_2021.pdf](https://www.dane.gov.co/files/investigaciones/condiciones_vida/pobreza/2021/Comunicado-pobreza-monetaria_2021.pdf).
3. Organización para la Cooperación y el Desarrollo Económicos (OCDE). (2023). "Cluster Analysis to Assess the Transferability of Public Health Interventions." Recuperado de [https://www.oecd.org/en/publications/cluster-analysis-to-assess-the-transferability-of-public-health-interventions\\_a5b1dcc1-en.html](https://www.oecd.org/en/publications/cluster-analysis-to-assess-the-transferability-of-public-health-interventions_a5b1dcc1-en.html).
4. National Institutes of Health (NIH). (2023). "Using Clustering for Socioeconomic Data in Public Health Studies." Publicado en *PMC*, Recuperado de <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9411746/>.
5. Papyrus Datos, Universidad de los Andes. "Panel de Salud y Servicios Públicos a Nivel Municipal." *Dataset disponible en* <https://papyrus-datos.co/dataset.xhtml?persistentId=doi:10.57924/RASU99>