

RECIBIDO EL 5 DE JULIO DE 2021 - ACEPTADO EL 6 DE OCTUBRE DE 2021

# CARACTERIZACIÓN DE LA FORMACIÓN DE LOS DOCENTES DE EDUCACIÓN BÁSICA Y MEDIA DE BOGOTÁ MEDIANTE ANÁLISIS ESPACIAL DE LOS DATOS

## CHARACTERIZATION OF THE TRAINING OF ELEMENTARY AND MIDDLE SCHOOL TEACHERS IN BOGOTÁ THROUGH SPATIAL DATA ANALYSIS

Gabriel Elías Chanchí Golondrino<sup>1</sup>

Manuel Alejandro Ospina Alarcón<sup>2</sup>

Martín Emilio Monroy Ríos<sup>3</sup>

### RESUMEN

Uno de los factores claves que inciden en

<sup>1</sup> Gabriel Elías Chanchí Golondrino. Doctor en Ingeniería Telemática - Universidad del Cauca. Profesor - Facultad de Ingeniería - Universidad de Cartagena. Colombiano, C.C. 10.301.274. gcchanchig@unicartagena.edu.co. ORCID: <https://orcid.org/0000-0002-0257-1988>.

<sup>2</sup> Manuel Alejandro Ospina Alarcón. Doctor en Ingeniería – Ciencia y Tecnología de Materiales – Universidad Nacional de Colombia -sede Medellín. Profesor - Facultad de Ingeniería - Universidad de Cartagena. Colombiano, C.C. 71.265.598. mospinaa@unicartagena.edu.co. ORCID: <https://orcid.org/0000-0003-4510-0753>.

<sup>3</sup> Martín Emilio Monroy Ríos. Doctor en Ingeniería Telemática – Universidad del Cauca. Profesor - Facultad de Ingeniería - Universidad de Cartagena. Colombiano, C.C. 80.411.302. mmonroyr@unicartagena.edu.co. ORCID: <https://orcid.org/0000-0003-4135-3251>.

la calidad de la educación básica y media es la formación profesional y de alto nivel de los docentes, aspecto sobre el cual no se ha evidenciado el desarrollo de estudios basados en análisis espacial de los datos. En este artículo se presenta como aporte la aplicación del análisis espacial de los datos en la caracterización de la formación profesional de los docentes de educación básica y media de la ciudad de Bogotá. El estudio propuesto en este artículo fue desarrollado sobre el dataset provisto por el

portal de datos abiertos de Bogotá para el año 2019, haciendo uso de la herramienta de software libre GeoDa, la cual permite realizar un análisis espacial de cuantiles y correlación lineal, así como aplicar de modelos de clustering mediante el uso del algoritmo K-Means. El estudio desarrollado, pretende servir de apoyo en cuanto a la toma de decisiones por parte de las autoridades pertinentes con respecto a la cobertura y la inversión en formación de los docentes en las diferentes localidades de la ciudad de Bogotá. Del mismo modo, este estudio pretende servir de referencia para ser extrapolado tanto a nivel de diferentes departamentos, como a nivel nacional, de cara a la caracterización y evaluación de la calidad de la educación en Colombia.

**PALABRAS CLAVES:** Análisis espacial, educación básica, educación media, formación docente, GeoDa.

## ABSTRACT

One of the key factors that have an impact on the quality of elementary and secondary education is the professional and high-level training of teachers, an aspect for which there is no evidence of studies based on spatial analysis of the data. In this article we present as a contribution the application of spatial analysis of data in the characterization of the professional training of elementary and middle school teachers in the city of Bogota. The study proposed in this article was developed from the dataset provided by the open data portal of Bogota for the year 2019, using the free software tool GeoDa, which allows a spatial analysis of quantiles and linear correlation, as well as the application of clustering models using

the K-Means algorithm. The developed study is intended to support decision making by the relevant authorities with respect to coverage and investment in teacher training in the different localities of the city of Bogota. Similarly, this study is intended to serve as a reference to be extrapolated both at the level of different departments, as well as at the national level, in order to characterize and evaluate the quality of education in Colombia.

**KEYWORDS:** Spatial analysis, elementary education, middle education, professional formation, GeoDa.

## 1. INTRODUCCIÓN

La calidad de la educación se ha convertido en un tema recurrente y en un desafío al que se enfrentan las instituciones educativas a nivel nacional e internacional, la cual puede definirse en función de tres aspectos fundamentales: pertinencia, eficacia y equidad. La calidad de la educación tiene influencia directa sobre la calidad de vida de un país, y no se logra únicamente con el establecimiento de políticas, aplicación de instrumentos o incremento en el financiamiento, sino a través de la formación de los docentes, de tal modo que a través de dicha formación se proporcione a estos actores las herramientas necesarias para afrontar los nuevos retos de la educación y la sociedad (Chehaybar, 2007; Martínez-Chairez et al., 2016; Miranda, 2007). La formación docente es un elemento indispensable para el desarrollo profesional y el mejoramiento de la práctica educativa, considerando las funciones y tareas que

realiza un docente dentro de su ámbito de trabajo (Pegalajar Palomino & Colmenero Ruíz, 2017; Sandoval, 2015).

En el contexto educativo, se ha evidenciado que la mayoría de los estudios de análisis de datos están enfocados en la aplicación de técnicas de análisis exploratorio y de minería de datos (Oviedo Carrascal & Jiménez Giraldo, 2019; Timarán-Pereira & Hidalgo-Troya, Arsenio Vidal-Alegría, 2020). Así mismo, en los diferentes repositorios de datos gubernamentales y en las estrategias de datos abiertos, son pocos los datasets que incluyen dentro de los atributos el componente geoespacial que permita la conducción de estudios de caracterización basados en análisis de datos espaciales. El análisis espacial de los datos comprende un conjunto de técnicas y modelos que utilizan específicamente la referencia espacial o geo-espacial de cada instancia u observación perteneciente al conjunto de datos, con el fin de determinar relaciones de tipo espacial (Goodchild & Haining, 2005; Rivero Sánchez, 2008; Sánchez-Peña, 2012). En la actualidad y en el contexto académico, una de las herramientas libres más difundidas para la conducción de este tipo de estudios es la herramienta GeoDa, la cual permite la aplicación de análisis de correlación lineal y espacial, el análisis a cuantiles sobre los datos asociados a un atributo del dataset y la aplicación de modelos de clustering mediante el uso del algoritmos de aprendizaje no supervisado (Anselin et al., 2010; Cerdán & Hernández, 2013; Leitner & Brecht, 2007; Yrigoyen, 2006).

En este artículo se propuso como aporte el desarrollo de un estudio de análisis espacial de los datos para la caracterización de la formación de los docentes de educación básica y media de las diferentes localidades de la ciudad de Bogotá. El estudio fue realizado sobre un dataset obtenido a partir del portal de datos abiertos de Bogotá, mediante la aplicación de las técnicas provistas por la herramienta de software libre GeoDa, la cual permite realizar a nivel espacial un análisis de correlación y de cuantiles, así como aplicar modelos de agrupamiento o clustering sobre la distribución de los datos a nivel espacial. Los resultados obtenidos a partir del estudio realizado, pretenden servir de referencia para la toma de decisiones por parte del Ministerio de Educación Nacional con respecto a la cobertura educativa y sobre los aspectos a mejorar a nivel de la inversión en la formación de los docentes de las diferentes localidades de la ciudad de Bogotá. Así mismo, este estudio busca ser extrapolado sobre los datos de otros departamentos y a nivel nacional con el fin de obtener indicadores de cara a mejorar la calidad de la educación en el país en los niveles básico y secundario.

El resto del artículo está organizado de la siguiente manera: en la sección 2 se presenta la metodología considerada para el desarrollo de la presente investigación; en la sección 3 se describen los resultados obtenidos en este trabajo, lo cual incluye la aplicación de las técnicas de análisis espacial de los datos mediante la herramienta GeoDa sobre el dataset con la formación de los docentes de educación básica y media para el año 2019; finalmente en la sección

4 se presentan las conclusiones y los trabajos futuros derivados de la presente investigación.

## 2. METODOLOGÍA

Para el desarrollo de la presente investigación, fueron definidas 4 fases metodológicas a

saber: selección y obtención de los datos, análisis de correlación, análisis de cuantiles y clustering, obtención de conclusiones sobre el estudio (ver Figura 1).



Figura 1. Metodología considerada

Dentro de la fase 1 de la metodología, fueron explorados un conjunto de datasets compatibles con la aplicación de las técnicas de análisis espacial y asociados a la formación profesional de docentes dentro de la ciudad de Bogotá, para lo cual se seleccionó a partir del portal de datos abiertos de Bogotá, un dataset con un total de 20 instancias correspondientes a los datos por localidad de la formación por diferentes niveles y género de los docentes. A partir del dataset seleccionado en la fase 1, dentro de la fase 2 se realizó el análisis de correlación entre los diferentes atributos del dataset, haciendo uso de la herramienta libre GeoDa y con el fin de determinar relaciones de dependencia lineal mediante la evaluación del coeficiente de determinación ( $R^2$ ). Dentro de la fase 3 de la metodología se realizó el análisis de la distribución espacial de los datos mediante la herramienta GeoDa, haciendo uso del método de segmentación por cuantiles y de la aplicación de modelos de agrupamiento o clustering a través del uso del algoritmo de aprendizaje no supervisado K-Means, haciendo uso de 5 clústeres y 1000 iteraciones. Finalmente, en

la fase 4 de la metodología, se consolidan los resultados del estudio y se obtienen un conjunto de conclusiones relacionadas con la distribución espacial de los datos de la formación de los docentes en las diferentes localidades de Bogotá.

## 3. RESULTADOS

En esta sección se presentan los resultados de la aplicación del análisis espacial de los datos sobre el dataset con la información de la formación de los docentes de las diferentes localidades de Bogotá para 2019, el cual fue obtenido del portal de datos abiertos de Bogotá. El dataset empleado consta de un total de 20 instancias y un total de 23 atributos asociados a la formación por niveles y por género de las localidades de Bogotá, así como los datos geo-espaciales de cada una de las instancias, los cuales posibilitan la representación visual del mapa en la herramienta GeoDa. De este modo, en la Tabla 1, son presentados los atributos principales del dataset, junto con su descripción y alcance.

Tabla 1. Descripción de los atributos del dataset de formación de docentes de Bogotá durante 2019

Atributo	Descripción
DOC_M	Número de docentes mujeres con título de doctor dentro de una localidad determinada
DOC_H	Número de docentes hombres con título de doctor dentro de una localidad determinada
MGS_M	Número de docentes mujeres con título de maestría dentro de una localidad determinada
MGS_H	Número de docentes hombres con título de maestría dentro de una localidad determinada
ESP_M	Número de docentes mujeres con título de especialización dentro de una localidad determinada
ESP_H	Número de docentes hombres con título de especialización dentro de una localidad determinada
PROF_M	Número de docentes mujeres con título profesional dentro de la localidad determinada
PROF_H	Número de docentes hombres con título profesional dentro de la localidad determinada
NORM_M	Número de docentes mujeres con formación de normalista dentro de una localidad determinada
NORM_H	Número de docentes hombres con formación de normalista dentro de una localidad determinada
BACH_M	Número de docentes mujeres con título de bachiller dentro de una localidad determinada
BACH_H	Número de docentes hombres con título de bachiller dentro de una localidad determinada
TEC_M	Número de docentes mujeres con título de tecnólogo dentro de una localidad determinada
TEC_H	Número de docentes hombres con título de tecnólogo dentro de una localidad determinada

A partir del dataset descrito en la Tabla 1, se realizó en primera instancia un análisis de correlación lineal comparando los atributos asociados a los géneros de los docentes, obteniendo que las correlaciones más altas se presentan entre las tres parejas de variables: PROF\_M y PROF\_H, MGS\_M y MGS\_H, ESP\_M, ESP\_H, con coeficientes

de correlación o determinación respectivos de 0.985, 0.982 y 0.979 (ver Figura 2). Del mismo modo, el anterior análisis permitió obtener que la correlación más baja entre los atributos se presenta en la pareja de variables: TEC\_M y TEC\_H, con un coeficiente de correlación o determinación de 0.001.

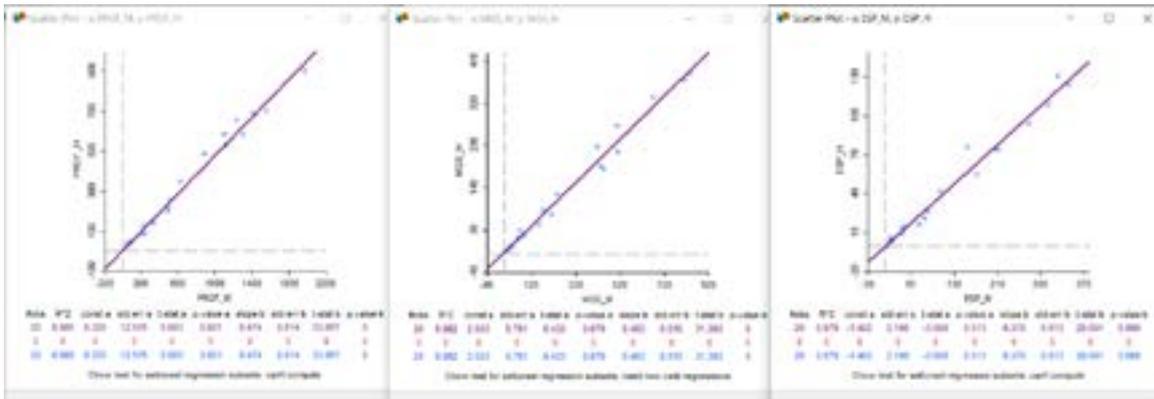


Figura 2. Mejores correlaciones obtenidas entre los atributos

De acuerdo a los anteriores resultados es posible concluir que las mayores correlaciones (coeficiente de determinación > 0.9) fueron obtenidas al comparar por géneros la formación profesional y de posgrado de los docentes, salvo a nivel de la formación de doctorado, en donde la correlación obtenida es baja, con un coeficiente de determinación de 0.184 (ver Figura 3).

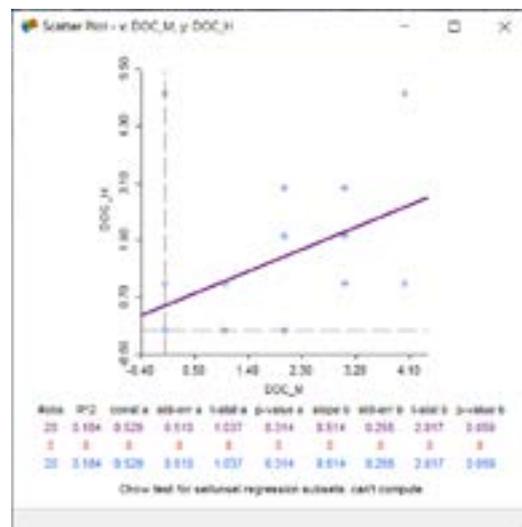


Figura 3. Correlación por género a nivel de formación doctoral

Una vez hecho el estudio de correlación, se procedió a realizar el análisis de cuantiles sobre los datos del dataset, el cual permite obtener grupos para un determinado atributo. Así en primera instancia, el la Figura 4 se presenta el análisis espacial a 5 cuantiles para los atributos asociados al

número de docentes mujeres con formación de posgrado en las localidades de Bogotá, es decir lo correspondiente a la formación de especialista (ESP\_M), magister (MGS\_M) y doctorado (DOC\_M).

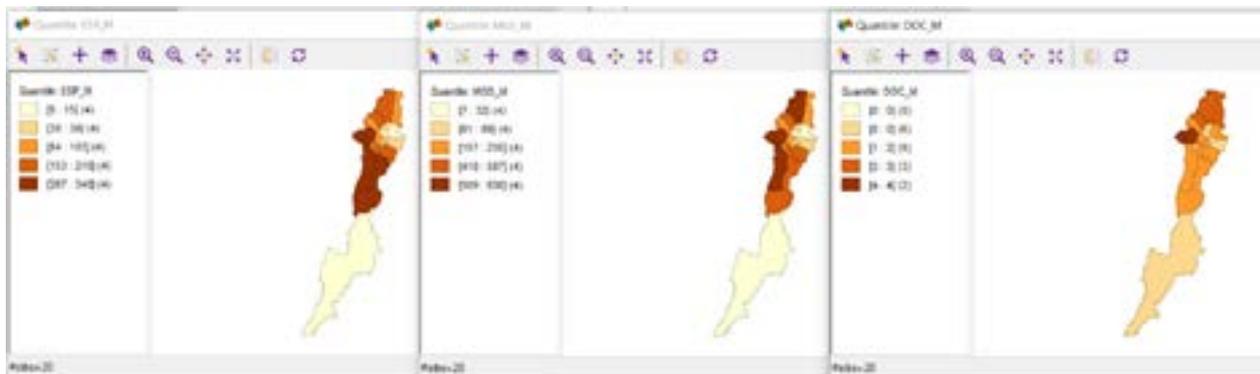


Figura 4. Análisis de cuantiles para docentes mujeres con formación de posgrado

De acuerdo a los resultados presentados en la Figura 4, es posible concluir que las localidades de: Sumapaz, Chapinero, Teusaquillo y La Candelaria, son las que cuentan con el menor número de docentes mujeres con formación a nivel de especialización. Así mismo, las cuatro localidades que cuentan con un mayor número de docentes mujeres con formación de especialización son: Ciudad Bolívar, Bosa, Kennedy y Usme. De otra parte, a nivel de la formación de maestría, es posible concluir que las localidades que tienen un menor número de docentes mujeres con este tipo de formación son: Sumapaz, Chapinero, Teusaquillo y la Candelaria, lo cual coincide con los resultados obtenidos a nivel de la formación a nivel de especialización. En este mismo sentido, en cuanto a las localidades que presentan un mayor número de docentes con formación de maestría se encuentran: Ciudad Bolívar,

Bosa, Kennedy y Suba, de tal modo que hay coincidencia en 3 localidades con respecto a los resultados obtenidos a nivel de especialización. Finalmente a nivel de la formación de doctorado en docentes mujeres, es posibles concluir que 6 de las 20 localidades no cuentan con docentes en este grado de formación (Sumapaz, Tunjuelito, Santafé, La Candelaria, Teusaquillo y Barrios Unidos). Así mismo, con respecto a las localidades que cuentan con un mayor número de docentes con formación de doctorado se encuentran: Bosa y Kennedy.

Continuando con el análisis a cuantiles, en la Figura 5 se presenta el análisis espacial a 5 cuantiles para los atributos asociados al número de docentes hombres con formación de posgrado en las localidades de Bogotá, es decir lo correspondiente a la formación de especialista (ESP\_H), magister (MGS\_H) y doctorado (DOC\_H).

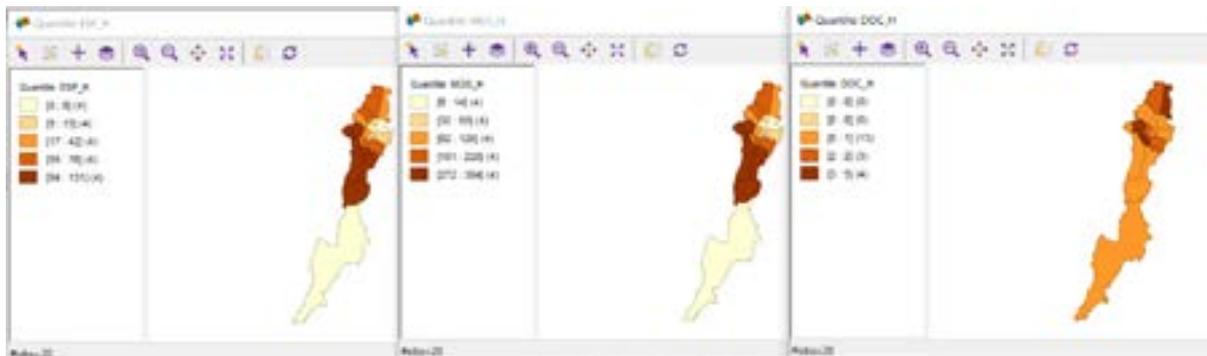


Figura 5. Análisis de cuantiles para docentes hombres con formación de posgrado

De acuerdo a los resultados presentados en la Figura 5, es posible concluir que las localidades de: Sumapaz, Chapinero, Teusaquillo y La Candelaria, son las que cuentan con el menor número de docentes hombres con formación de especialización. Del mismo modo, las cuatro localidades que tienen un mayor número de docentes hombres con formación de especialización son: Ciudad Bolívar, Bosa, Kennedy y Usme. Se puede observar que las localidades que cuentan con el menor y el mayor número de docentes hombres a nivel de especialización son las mismas que se obtuvieron para el caso de las docentes mujeres. De otra parte, a nivel de la formación de maestría, es posible concluir que las localidades que tienen un menor número de docentes hombres con este tipo de formación son: Sumapaz, Chapinero, Teusaquillo y La Candelaria, lo cual coincide con los resultados obtenidos a nivel de la formación de especialización. En ese mismo sentido, en lo referente a las localidades

que cuentan con un mayor número de docentes con formación de maestría están: Ciudad Bolívar, Bosa, Kennedy y Usme, de tal manera que hay concordancia con las localidades con mayor número de docentes a nivel de la especialización. Es posible observar que las localidades que cuentan con el menor número de docentes hombres a nivel de maestría, son las mismas que se obtuvieron para el caso de las mujeres, mientras que en el caso de las localidades que cuentan con el mayor número de docentes hombres a nivel de maestría, tres de ellas también fueron obtenidas para el caso de las mujeres (Ciudad Bolívar, Bosa y Kennedy).

Siguiendo con el análisis a cuantiles, en la Figura 6 es presentado el análisis a 5 cuantiles para los atributos asociados al número de docentes mujeres con formación profesional (PROF\_M), técnica (TEC\_M) y de normalista (NORM\_M) en las localidades de Bogotá.

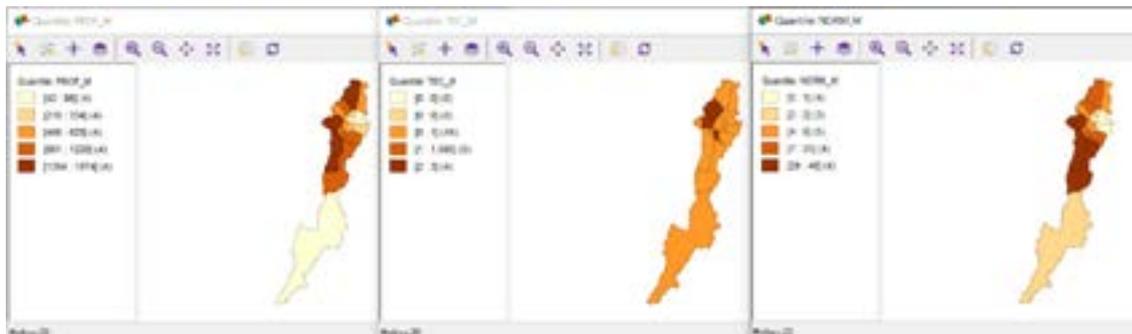


Figura 6. Análisis de cuantiles para docentes mujeres con formación profesional, técnica y normalista

Según los resultados presentados en la Figura 6, es posible concluir que las localidades que cuentan con un menor número de docentes mujeres con formación profesional son: Sumapaz, Chapinero, Teusaquillo y La Candelaria. De la misma manera, las cuatro localidades que cuentan con el mayor número de docentes mujeres con formación profesional son: Ciudad Bolívar, Bosa, Kennedy y Suba. Se puede observar que las localidades que cuentan con el menor número de docentes mujeres con formación profesional, son las mismas obtenidas para la formación de especialización y de maestría. Por otro lado, con respecto a las localidades que tienen un mayor número de docentes mujeres con formación profesional, son las mismas obtenidas para la formación de maestría. De otra parte, a nivel de la formación técnica, es posible concluir que las 4 localidades que tienen un mayor número de docentes mujeres con este tipo de formación son: Rafael Uribe, Kennedy, Fontibón y Engativá, mientras que las restantes 16 localidades tienen entre 0 y 1 docentes mujeres con formación técnica. Del mismo modo, a nivel de las localidades que tienen un menor número de docentes mujeres con formación

de normalista se encuentran: Santafé, Chapinero, Teusaquillo, La Candelaria, de tal modo que hay coincidencia en tres localidades (Chapinero, Teusaquillo, La Candelaria) con respecto a los resultados obtenidos en las localidades con menor número de docentes mujeres a nivel de la formación profesional. Por otra parte, las 4 localidades que tienen un mayor número de docentes mujeres con formación de normalista son: Ciudad Bolívar, Bosa y Kennedy y Usme, de tal manera que hay coincidencia en 3 localidades (Ciudad Bolívar, Bosa y Kennedy) con respecto a las asociadas a las de mayor número de docentes mujeres con formación de profesionales.

Continuando con el análisis de cuantiles, en la Figura 7 se presenta el análisis a 5 cuantiles para los atributos asociados al número de docentes hombres con formación profesional (PROF\_H), técnica (TEC\_H) y de normalista (NORM\_H) en las diferentes localidades de Bogotá.

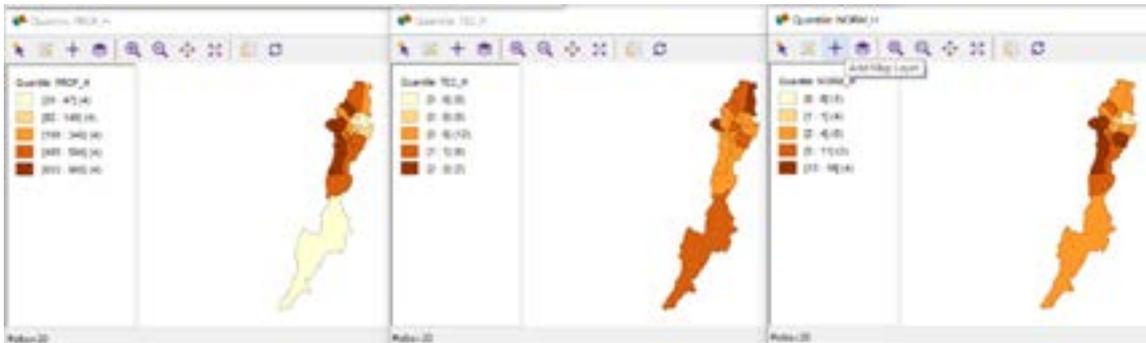


Figura 7. Análisis de cuantiles para docentes hombres con formación profesional, técnica y normalista

Según los resultados presentados en la Figura 7, es posible concluir que las localidades que cuentan con un menor número de docentes hombres con formación profesional son: Sumapaz, Chapinero, Teusaquillo y La Candelaria. De la misma manera, las cuatro localidades que cuentan con el mayor número de docentes hombres con formación profesional son: Ciudad Bolívar, Bosa, Kennedy y Engativá. Es posible observar que las localidades que cuentan con el menor número de docentes hombres con formación profesional, son las mismas obtenidas en cuanto a la formación de docentes hombres en la formación de especialización y de maestría.

Por otro lado, con respecto a las localidades que tienen un mayor número de docentes hombres con formación profesional, se evidencia que 3 de estas localidades aparecen también en los resultados de la formación de docentes hombres a nivel de especialización y maestría.

De otra parte, en lo referente a la formación técnica de docentes hombres, se obtuvo que 6 localidades (Sumapaz, San Cristóbal, Puente Aranda, Los Mártires, Engativá y

Suba) cuentan cada una con un docente con formación técnica, mientras que 2 localidades (Bosa y Usaquen) cuentan cada una con dos docentes con formación técnica. Del mismo modo, con respecto a la formación normalista de los docentes hombres, existe una localidad (Chapinero) que no cuenta con ningún docente normalista, mientras que existen 4 localidades (Antonio Nariño, Los Mártires, Teusaquillo y La Candelaria) que cuentan cada una con un docente normalista, de tal modo que dos de estas localidades (Teusaquillo y la Candelaria) también fueron obtenidas en el análisis del menor número de docentes mujeres normalistas. Así mismo, las 4 localidades que cuentan con un mayor número de docentes hombres normalistas son: Ciudad Bolívar, San Cristóbal, Bosa y Kennedy, de tal manera que tres de estas localidades (Ciudad Bolívar, Bosa y Kennedy) también se obtuvieron en el análisis del mayor número de docentes mujeres normalistas.

Una vez fue realizado el análisis a cuantiles sobre el dataset, a continuación, se procedió a realizar el análisis de clustering o agrupamiento, mediante el uso del algoritmo de aprendizaje no supervisado K-Means,

el cual permite obtener la distribución espacial de los datos en un conjunto de clústeres, cada uno con un centroide asociado, alrededor del cual se concentran los datos de cada clúster (Chanchí et al., 2020; Likas et al., 2003). Así, en la Figura 8 se presentan los resultados del análisis

de agrupamiento a 5 clústeres sobre los atributos asociados al número de docentes mujeres con formación de posgrado en las localidades de Bogotá, lo cual corresponde a la relación en el espacio de los atributos: DOC\_M, MGS\_M, ESP\_M.

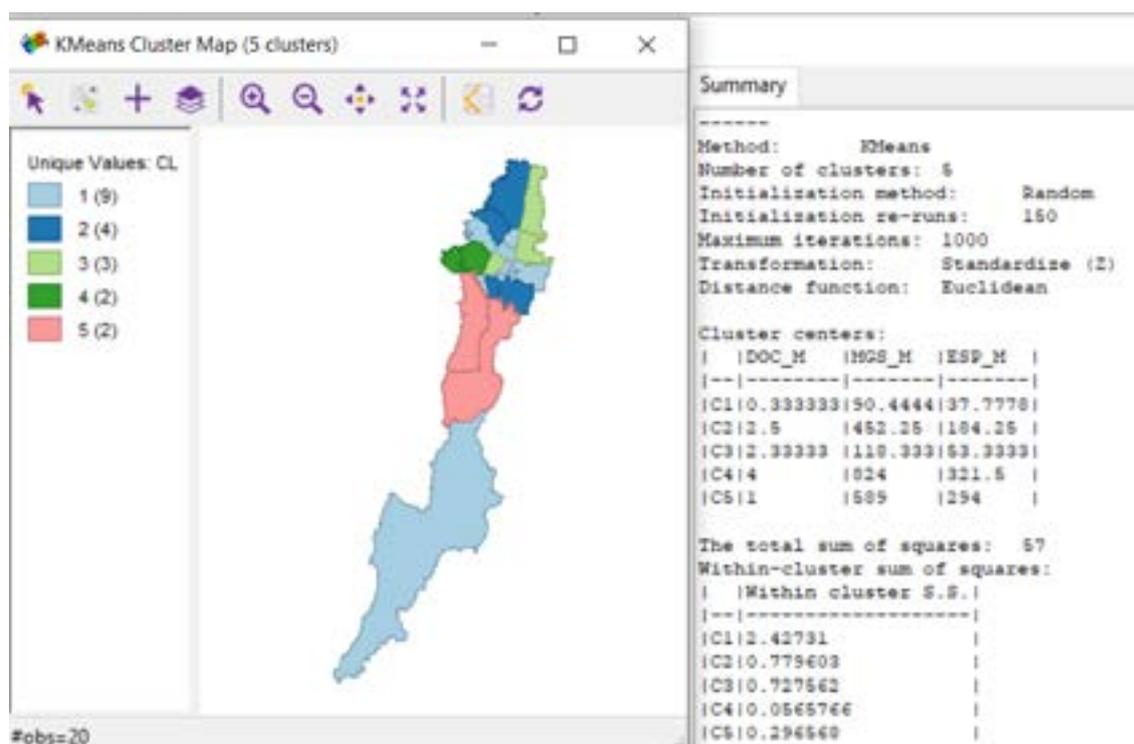


Figura 8. Análisis de clustering para docentes mujeres con formación de posgrado

Se puede observar en la Figura 8 como el clúster que cuenta con más instancias o localidades asociadas es el clúster 1, cuyo centroide es  $C1 = \{DOC\_M = 0.333, MGS\_M = 90.444, ESP\_M = 37.778\}$ . De este modo, el clúster 1 agrupa geográficamente a 9 localidades a saber: Sumapaz, Tunjuelito, Antonio Nariño, Los Mártires, La Candelaria, Santafé, Teusaquillo, Fontibón y Barrios Unidos. Lo anterior implica que el 45% de las localidades de Bogotá tienen los atributos DOC\_M, MGS\_M y ESP\_M concentrados alrededor del centroide C1.

Por otra parte el clúster que agrupa dos instancias y/o localidades (Bosa y Kennedy) y que contiene los valores mayores de los DOC\_M, MGS\_M y ESP\_M es el centroide  $C4 = \{DOC\_M = 4, MGS\_M = 824 \text{ y } ESP\_M = 21.5\}$ . Lo anterior quiere decir que el 10% de las localidades de Bogotá concentra el mayor número de docentes mujeres con una mejor formación especializada o formación a nivel de doctorado, maestría y especialización. Del mismo modo, en la Figura 9 son presentados los resultados del análisis de agrupamiento a 5 clusters

sobre los atributos asociados al número de docentes hombres con formación a nivel de posgrado dentro de las localidades de Bogotá, lo cual corresponde a la relación espacial de los atributos: DOC\_H, MGS\_H y ESP\_H.

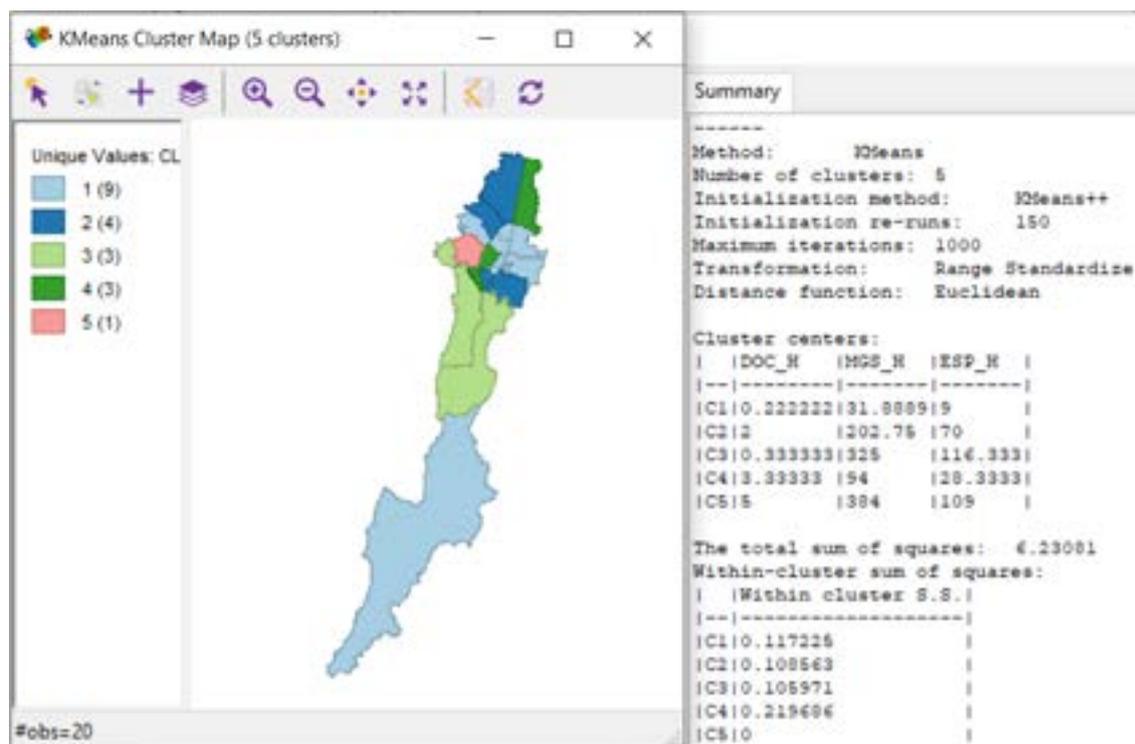


Figura 9. Análisis de clustering para docentes hombres con formación de posgrado

Es posible observar en la Figura 9 como el clúster que cuenta con un mayor número de instancias o localidades asociadas es el clúster 1, cuyo centroide es  $C1 = \{DOC\_H=0.222, MGS\_H=31.889, ESP\_H=9\}$ . De esta manera, el clúster 1 agrupa geográficamente a 9 localidades a saber: Sumapaz, Antonio Nariño, Los Mártires, La Candelaria, Santafé, Fontibón, Teusaquillo, Barrios Unidos y Chapinero. Lo anterior implica que el 45% de las localidades de Bogotá tienen los atributos DOC\_H, MGS\_H y ESP\_H concentrados alrededor del

centroide C1. De otra parte, el clúster que agrupa una instancia o localidad (Kennedy) y contiene los valores mayores de los atributos: DOC\_H, MGS\_H y ESP\_H es el centroide  $C5 = \{DOC\_H=5, MGS\_H=3.84, ESP\_H=109\}$ . Lo anterior quiere decir que el 5% de las localidades de Bogotá concentra el mayor número de docentes hombres con una mejor formación especializada o formación a nivel de doctorado, maestría y especialización. Se puede observar que la localidad de Kennedy apareció tanto en los resultados para las docentes mujeres

como para los docentes hombres. En concordancia con lo anterior, si se aplican los modelos de clustering a las parejas de atributos DOC\_M y DOC\_H, MGS\_M y MGS\_H, se obtiene como resultado las dos gráficas de distribución espacial mostradas en la Figura 10.

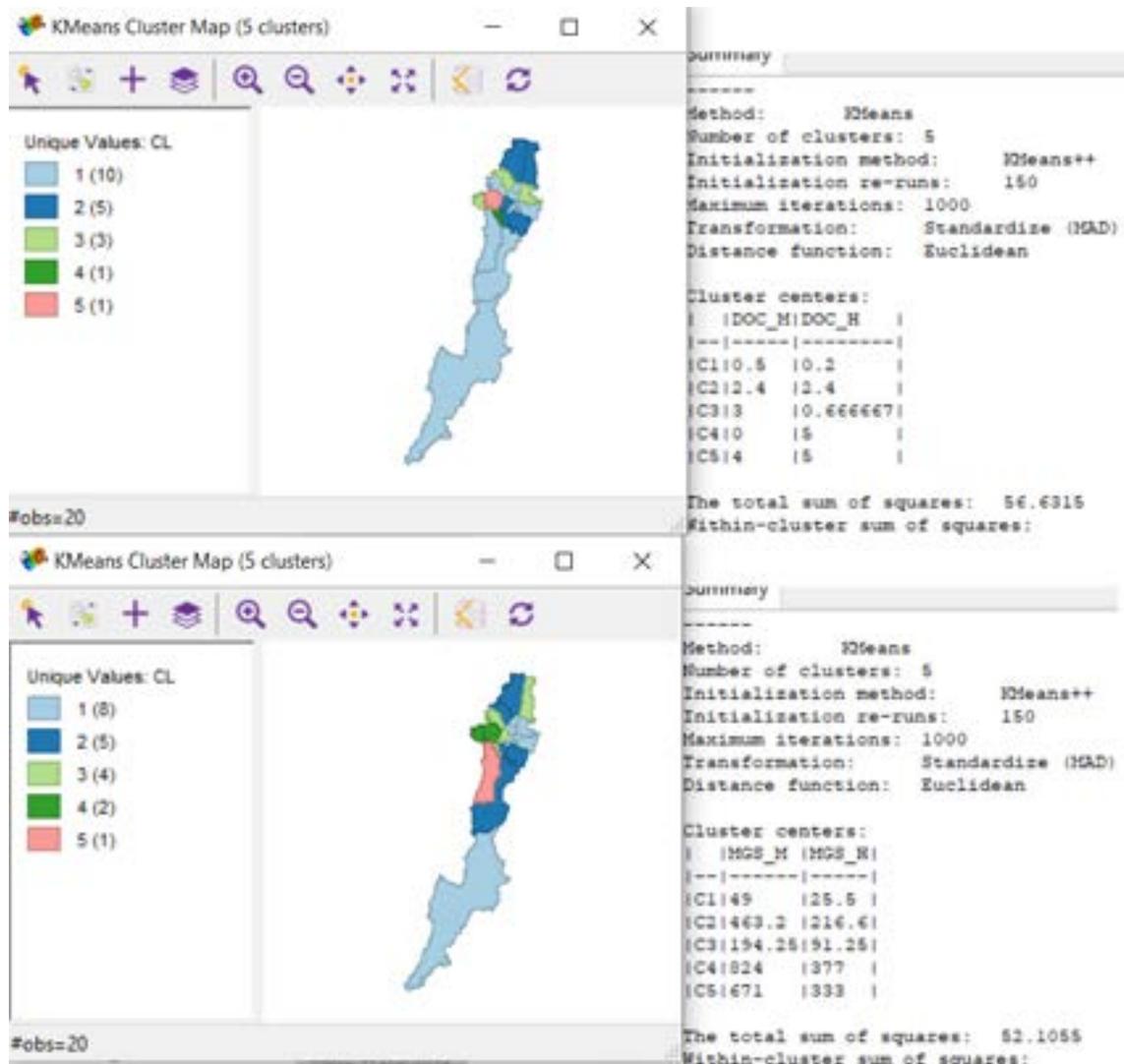


Figura 10. Análisis de clustering para docentes hombres y mujeres con formación de maestría y doctorado

En la gráfica superior de la Figura 10, se observa como al relacionar los atributos DOC\_M y DOC\_H, se obtiene que el clúster que presenta los mayores valores de estos atributos es el clúster 5, el cual agrupa la localidad de Kennedy y cuyo centroide es  $C5=\{DOC\_M=4, DOC\_H=5\}$ . Esto quiere decir que la localidad de Kennedy cuenta de manera conjunta con el mayor número de docentes hombre y mujeres a nivel de maestría y doctorado. Del mismo modo, en la gráfica inferior de la Figura 10, se puede observar como al relacionar los atributos MGS\_M y MGS\_H, se obtiene que el clúster que presenta los mayores valores de estos atributos es el clúster 4, el cual tiene por centroide  $C4=\{MGS\_M=824 \text{ y } MGS\_H=377\}$  y está distribuido geográficamente en las localidades de: Bosa y Kennedy. Esto quiere decir que estas localidades cuentan

de manera conjunta con el mayor número de docentes hombres y mujeres con formación de maestría. Del mismo modo, el clúster que agrupa los menores valores de los atributos MGS\_M y MGS\_H es el clúster 1, cuyo centroide es  $C1=\{MGS\_M=49, MGS\_H=25.5\}$  y cuyas localidades asociadas son: Sumapaz, Antonio Nariño, Los Mártires, La Candelaria, Santafé, Teusaquillo, Barrios Unidos y Chapinero. Lo anterior quiere decir que las localidades mencionadas, son las que cuentan con un menor número de docentes hombres y mujeres con formación de maestría. De otra parte, en la Figura 11 se presentan los resultados de la aplicación del modelo de clustering para los atributos NORM\_M, BACH\_M y TEC\_M, lo cual involucra la distribución espacial de las docentes mujeres con formación de normalistas, bachilleres y técnicas.

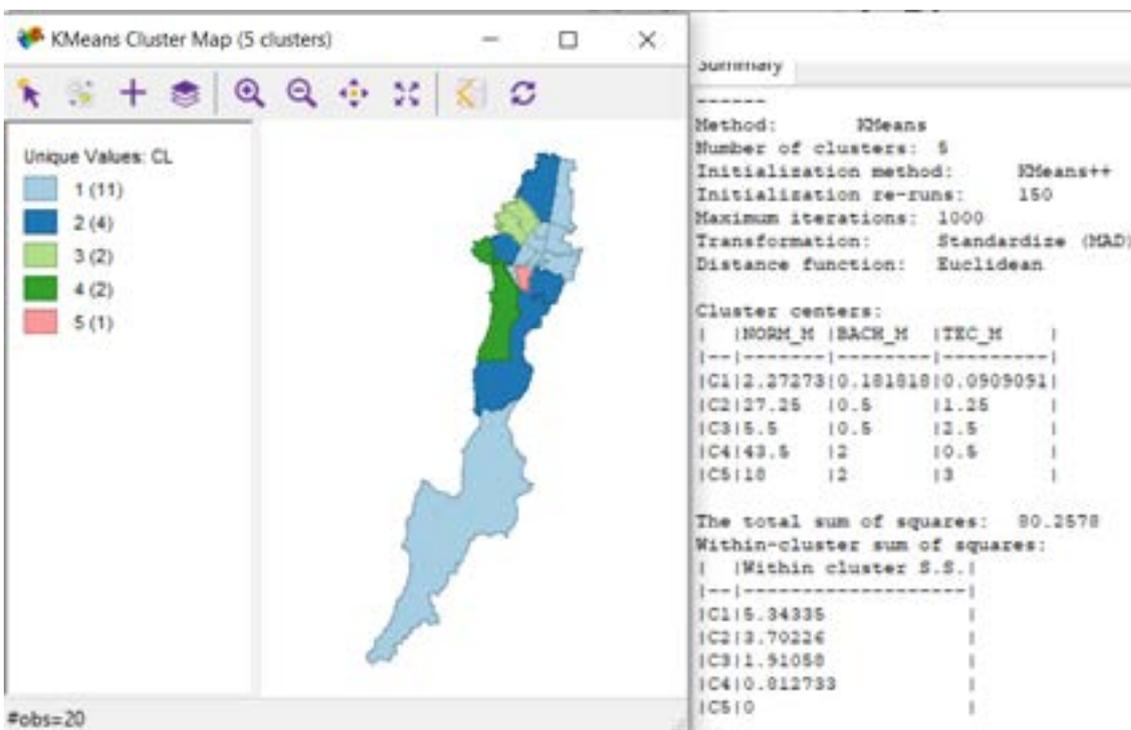


Figura 11. Análisis de clustering para docentes mujeres con formación de normalistas, bachilleres y técnicas

De acuerdo a la Figura 11, es posible observar como el clúster que cuenta con los mayores valores de los atributos NORM\_M, BACH\_ y TEC\_M es el clúster 5, cuyo centroide es  $C5=\{NORM\_M=18, BACH\_M=2 \text{ y } TEC\_M=3\}$  y cuya localidad asociada es: Rafael Uribe. Lo anterior quiere decir que esta localidad es la que cuenta con un mayor número de docentes mujeres con formación de normalistas, bachilleres y técnicas. Finalmente, en la Figura 12 se presentan los resultados de la aplicación de los modelos de clustering

para los atributos NORM\_H, BACH\_H y TEC\_H, lo cual involucra la distribución espacial de los docentes hombres con formación de normalistas, bachilleres y técnicos. De acuerdo a la Figura 12, es posible observar como el clúster que cuenta con los valores de los atributos NORM\_H, BACH\_ y TEC\_M diferentes de 0 es el clúster 3, cuyo centroide es  $C3=\{NORM\_H=13.667, BACH\_H=1 \text{ y } TEC\_H=0.333\}$  y cuyas localidades asociadas son: Ciudad Bolívar, Kennedy y San Cristóbal.

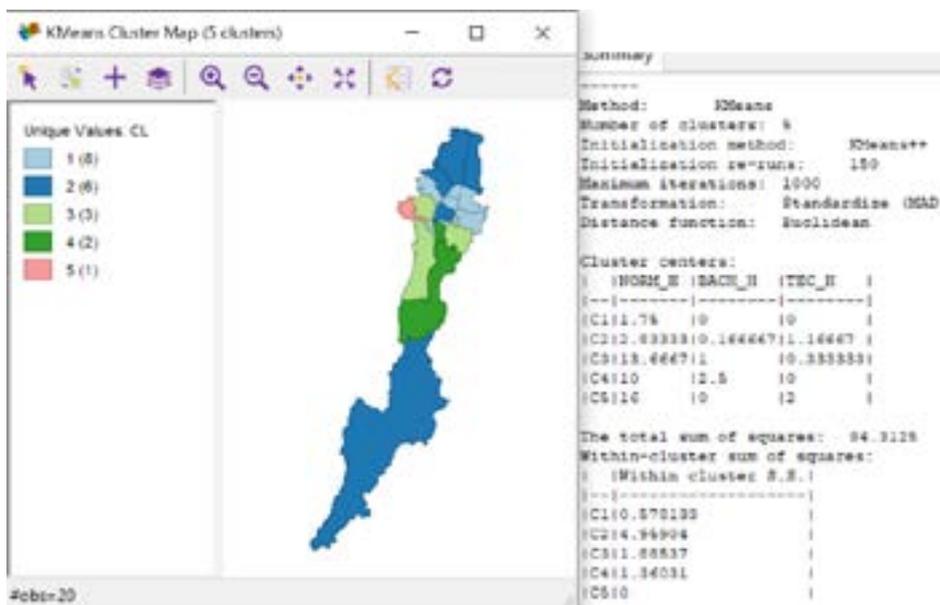


Figura 12. Análisis de clustering para docentes mujeres con formación de normalistas, bachilleres y técnicas

## 5. CONCLUSIONES Y TRABAJOS FUTUROS

Teniendo en cuenta que la mayoría de los estudios realizados sobre datos en el contexto educativo han sido realizados mediante la aplicación de técnicas de minería de datos, en este artículo se presentó como aporte el desarrollo de un estudio basado en el análisis espacial de

los datos, sobre el dataset de formación de docentes en las 20 localidades de Bogotá para el año 2019, el cual fue obtenido a partir del portal de datos abiertos de Bogotá. El estudio desarrollado pretende servir de referencia para la conducción de estudios similares de análisis espacial de los datos en otras regiones del país, de cara a la toma de decisiones por parte de las autoridades

pertinentes con respecto a la cobertura y la calidad educativa.

Para el desarrollo de estudios de análisis espacial de los datos, no solo es importante contar con los diferentes atributos a relacionar, sino con los atributos adicionales que involucran el componente geoespacial y permiten la visualización de los datos mediante mapas en las herramientas de análisis espacial de los datos como GeoDa. En este sentido, el dataset empleado contaba tanto con los atributos geoespaciales asociados a las 20 localidades de Bogotá, como los atributos correspondientes al número de docentes hombres y mujeres con formación de posgrado, profesional, técnica y normalista. De este modo el dataset empleado está formado por 20 instancias y un total de 24 atributos.

Para la conducción del estudio presentado en este artículo, se hizo uso de la herramienta libre de análisis espacial de los datos GeoDa, la cual permitió realizar un análisis de correlación lineal entre los atributos del dataset, un análisis a cuantiles sobre los atributos individuales y finalmente un análisis de clustering relacionando diferentes atributos mediante el uso del algoritmo de aprendizaje no supervisado K-Means. A partir de los resultados obtenidos mediante el uso de la herramienta GeoDa, fue posible obtener las localidades de Bogotá, en donde es necesario fortalecer la formación de los docentes de educación básica y media, de cara a mejorar la calidad de la educación.

Como trabajo futuro derivado del presente artículo, se pretende retroalimentar el estudio desarrollado, mediante la aplicación

de técnicas de minería de datos sobre los atributos del dataset. En este mismo sentido, se pretende extrapolar el presente estudio a otros municipios y departamentos de Colombia, con el fin de facilitar la toma de decisiones a nivel de planeación sobre la cobertura educativa y la formación de los docentes.

## AGRADECIMIENTOS

Los autores del artículo agradecen a la Universidad de Cartagena por el apoyo brindado en el desarrollo de la presente investigación.

## REFERENCIAS BIBLIOGRÁFICAS

- Anselin, L., Syabri, I., & Kho, Y. (2010). GeoDa: An Introduction to Spatial Data Analysis. In *Handbook of Applied Spatial Analysis*. Springer, Berlin, Heidelberg. [https://doi.org/10.1007/978-3-642-03647-7\\_5](https://doi.org/10.1007/978-3-642-03647-7_5)
- Cerdán, F. T., & Hernández, F. A. L. (2013). Distribución espacial de la actividad económica en la comarca de Cartagena. Un análisis basado en la demografía empresarial usando Arcview y Geoda. *Rect@*, 4, 97–118.
- Chanchí, G., Ospina, M., & Pérez, J. (2020). Sistema IoT para la monitorización de la variabilidad del ritmo cardiaco en pruebas de usabilidad. *Revista Espacios*, 41(25), 84–97. <http://es.revistaespacios.com/a20v41n25/a20v41n25p07.pdf>
- Chehaybar, E. (2007). Reflexiones sobre el papel del docente en la calidad educativa. *Reencuentro. Análisis De*

- Problemas Universitarios*, 50, 100–106. <https://reencuentro.xoc.uam.mx/index.php/reencuentro/article/view/639>
- Goodchild, M., & Haining, R. (2005). SIG y análisis espacial de datos: perspectivas convergentes. *Investigaciones Regionales - Journal of Regional Research*, 6, 175–201.
- Leitner, M., & Brecht, H. (2007). Software Review: Crime Analysis and Mapping with GeoDa 0.9.5-i. *Social Science Computer Review*, 25(2), 265–271. <https://doi.org/10.1177/0894439307298921>
- Likas, A., Vlassis, N., & J. Verbeek, J. (2003). The global k-means clustering algorithm. *Pattern Recognition*, 36(2), 451–461. [https://doi.org/10.1016/S0031-3203\(02\)00060-2](https://doi.org/10.1016/S0031-3203(02)00060-2)
- Martínez-Chairez, G. I., Guevara-Araiza, A., & Valles-Ornelas, M. (2016). El desempeño del docente y la calidad educativa. *Ra Ximhai*, 12(6), 123–134.
- Miranda, C. (2007). Educación Superior, Mecanismos de Aseguramiento de la Calidad y Formación Docente. *Estudios Pedagógicos*, 1, 95–108. <https://doi.org/http://dx.doi.org/10.4067/S0718-07052007000100006>
- Oviedo Carrascal, A. I., & Jiménez Giraldo, J. (2019). Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas SABER-PRO. *Revista Politécnica*, 15(29), 128–140. <https://revistas.elpoli.edu.co/index.php/pol/article/view/1499/1219>
- Pegalajar Palomino, M. del C., & Colmenero Ruíz, M. de J. (2017). Actitudes y formación docente hacia la inclusión en Educación Secundaria Obligatoria. *Revista Electrónica de Investigación Educativa*, 19(1), 84–97. <http://www.scielo.org.mx/pdf/redie/v19n1/1607-4041-redie-19-01-00084.pdf>
- Rivero Sánchez, M. (2008). Análisis espacial de datos y Turismo : Nuevas técnicas para el análisis turístico. *Revista de Estudios Empresariales*, 2, 48–66.
- Sánchez-Peña, L. L. (2012). Alcances y límites de los métodos de análisis espacial para el estudio de la pobreza urbana. *Papeles de Poblacion*, 18(72), 147–179.
- Sandoval, F. (2015). La formación permanente del docente en Latinoamérica. *Revista Iberoamericana Para La Investigación y El Desarrollo Educativo*, 6(11). <https://www.redalyc.org/pdf/4981/498150319006.pdf>
- Timarán-Pereira, R., & Hidalgo-Troya, Arsenio Vidal-Alegría, F. (2020). Una Mirada al Desempeño Académico en las Pruebas Saber Pro de los Estudiantes de Ingeniería desde la Minería de Datos Educativa. *Revista Ibérica de Sistemas e Tecnologías de Informação*, E38, 29–42.
- Yrigoyen, C. C. (2006). Análisis estadístico de datos geográficos en geomarketing: el programa GeoDa. *Distribución y Consumo*, 178, 34–45.