

ISAE



ALCALDÍA MAYOR
DE BOGOTÁ D.C.

SECRETARÍA DE
PLANEACIÓN



ALCALDÍA MAYOR DE BOGOTÁ
SECRETARIA DISTRITAL DE PLANEACIÓN - SDP

Claudia Nayibe López Hernández
Alcaldesa Mayor de Bogotá D.C.

Felipe Jiménez Ángel
Secretario Distrital de Planeación

Andrés Leonardo Acosta Hernández
Subsecretario de Información

Juan David Vinasco Idárraga
Director de Información y Estadísticas

Diego Felipe López Ospina
Investigador

**Prospectiva de empleo a nivel territorial:
Indicador de Seguimiento a la Actividad Económica**

Diciembre 2023



INTRODUCCIÓN

En el marco del Plan de Ordenamiento Territorial adoptado en el año 2021, la administración distrital dispuso como uno de los indicadores de seguimiento al mismo el Índice de Seguimiento a la Actividad Económica (ISAE).

En aras de robustecer el análisis y realizar el primer seguimiento del mismo, se dispuso realizar una estimación del ISAE para el año 2021 y 2023 de tal forma que se cumpla con la disposición del POT.

El presente estudio profundiza los análisis del ISAE, incorporando además del componente estadístico, análisis espaciales y estimación de variación a nivel de manzana, localidad y UPL. De otro lado, el estudio se enmarca en la línea de estudios de la Dirección de Información y Estadísticas asociada a prospectiva de empleo al realizar las variaciones entre los años 2021 y 2023.



1. DATOS Y METODOLOGÍA

La información que se utiliza para el desarrollo del ejercicio parte de dos fuentes, la primera la información de establecimientos de comercio de la Cámara de Comercio de Bogotá (CCB) mientras que la segunda corresponde a la información de construcciones por destino económico de la Unidad Administrativa Especial de Catastro Distrital (UAECD).

Dado que el análisis se hace de forma intertemporal, partiendo del momento inicial que es la expedición del Decreto Distrital 555 de 2021 – Plan de Ordenamiento Territorial, de tal forma que de la fuente de CCB se tendrá en cuenta el número de establecimientos de comercio y empleo reportado para aquellas observaciones que estén activas con al menos 3 años de renovación de la matrícula mercantil. Dicha información se encuentra geolocalizada según la dirección del registro, proceso que realiza la SDP. Para los análisis se tendrá en cuenta los puntos de corte a marzo del año 2021 y marzo del año 2023 para el procesamiento de la información.

En relación con los datos de construcciones, se realiza una agrupación de los destinos económicos, de tal forma que correspondan a tres grandes categorías: comercio, servicio e industria. Los datos de la UAECD procesados se tomaron para los años 2021 y 2023, agrupándolos a nivel de manzana catastral al igual que el caso de la información de CCB.

Para agregar la información se plantea utilizar métodos que permitan establecer pesos de las variables para la composición de un índice que sean exógenas y correspondan a la variabilidad de la información.

Los métodos de análisis multivariados en estadística buscan dar información respecto a dos clases de alcance, a saber: interdependencia de las variables y dependencia entre una o varias variables respecto a otra u otras. Respecto al análisis

por componentes principales, Díaz y Morales (2012) explican que este método busca la “(...) estructuración de un conjunto de datos multivariados mediante la reducción del número de variables”. De esta manera, se transforma un conjunto de variables base en un conjunto menor de variables las cuales “contienen la mayor parte de la variabilidad presente del conjunto inicial” o base.

El análisis de componentes principales (ACP) tiene como objetivo originar nuevas variables que expresen la información contenida en un conjunto de datos base, reducir la dimensión del espacio dónde están los datos, eliminar en lo posible las variables que aporten poco y facilitar la interpretación de la información contenida en los datos base (Díaz y Morales, 2012).

A nivel general, el ACP se deriva de una matriz Y que cuenta con dimensiones n x p en donde p representa el número de variables identificadas y n es el número de individuos encuestados.

$$\mathbf{Y} = \begin{matrix} & \begin{matrix} X_1 & X_2 & \dots & X_p \end{matrix} \\ \begin{matrix} I_1 \\ I_2 \\ I_r \\ \dots \\ I_n \end{matrix} & \begin{vmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ x_{r1} & x_{r2} & \dots & x_{rp} \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{vmatrix} \end{matrix}$$

El número de variables p se define de acuerdo con las variables a trabajar, que para el caso son 5, a saber: metros cuadrados construidos en comercio, servicios, industria, número de establecimientos de comercio y total de empleos reportados.

El número de individuos n corresponde al total manzanas censales. Para dar origen a la matriz Y se estandarizan sus valores y de esta forma contar con varianza unitaria

y promedios nulos para cada variable, utilizando para esto el método de distancia del mejor y peor (Freudenberg, 2003).

$$x_{np} = \left[\frac{\text{valor actual} - \text{valor mínimo}}{\text{valor máximo} - \text{valor mínimo}} \right] 100 \quad (1)$$

El dominio de cada valor de $x_{np} \in [0,1]$. Con ello se garantiza contar con valores entre 0 y 1.

La posibilidad de la aplicación del método implica superar algunas pruebas que permiten considerar el ACP como el apropiado para su desarrollo. La prueba de Barlett es una de ellas, pues determina si el conjunto de variables seleccionadas son homogéneas en varianza; una segunda prueba es el criterio Kaiser-Mayer-Olkin, esta prueba permite determinar qué tan ajustada está la información para ser utilizada en un ejercicio de análisis factorial. De acuerdo con el resultado, la información está en menor o mayor medida adecuada según el siguiente criterio: entre 0.5 y 0.7 es mediocre, entre 0.7 y 0.8 es bueno, 0.8 a 0.9 es muy bueno y entre 0.9 y 1 es excelente.

Finalmente se estiman los valores propios y las cargas por cada componente; del primero se identifica la varianza de las nuevas componentes y del segundo el peso de cada variable por cada componente. Ambos valores son relevantes pues permiten determinar la cantidad de componentes a utilizar en función de la varianza, mientras que el peso o carga de las variables son utilizadas para la construcción del índice. De forma general el índice para cada unidad espacial es la siguiente:

$$y_{in} = \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \dots + \alpha_n x_{in} \quad (2)$$

En donde y_{in} corresponde al índice estimado para la manzana censal i , α es la carga (peso) derivado del componente y x_i es la variable que forma parte de la

conformación del índice. Es importante indicar que para la construcción de un único índice se requiere que la varianza explicada supere al menos un 70%, por ello se puede requerir más de un componente para cumplimiento y, por ende, la forma general explicitada debe ser replicada por cada componente seleccionado.

Una vez calculado el ISAE por componentes principales es necesario espacializar los resultados para cada manzana y conocer los patrones en función de la actividad económica calculada.

Por ello se recurre a una metodología que permita expandir los datos en forma patrones no areales en el espacio. Para ello se utiliza el análisis de densidad *kernel*, el cual se realiza a nivel de celda y tiene en cuenta la cantidad de nodos (para el caso centroides de las manzanas catastrales) y el valor propio del atributo a expandir espacialmente, el cual es el ISAE que se calcule para los años 2021 y 2023. De forma general siguiendo a Silverman (1998) así como a Connor y Connor (2009) en la ecuación 1 se podrá analizar la formalización de la estimación del análisis de densidad kernel.

$$Densidad = \frac{1}{radio^2} \sum_{i=1}^n \left[\frac{3}{\pi} \cdot ISAE_i \left(1 - \left(\frac{Dist_i}{radio} \right)^2 \right)^2 \right] \quad (3)$$

Las distancias euclidianas calculadas entre los nodos se realizan por el método planar dada la escala de análisis (ciudad) utilizado en el ejercicio. El ancho de banda o radio se basa en la identificación de datos atípicos de Silverman (1996). El software utilizado para la estimación es ArcGIS de ESRI.

Para utilizar la información areal o de polígonos, que son las delimitaciones de las manzanas catastrales, se hace uso de los *Local Indicators of Spatial Association – LISA* presentados por Anselin (1995). Con ello se logra identificar la presencia de aglomeraciones espaciales (*cluster*) o datos atípicos espaciales. La metodología

parte de la construcción de una matriz de pesos espaciales que, para el caso, se realiza a partir de la distancia euclidiana en metros entre los más de 39 mil nodos, que a su vez es estandarizada por filas. La ecuación 4 contiene la forma de cálculo de cada indicador local.

$$I_i = \frac{\sum_j w_{ij} z_i z_j}{\sum_i z_i^2} \quad (4)$$

En donde w_{ij} es la matriz de pesos espaciales mientras que los valores $z_i z_j$ son desviaciones de la media, en tanto que la suma de J corresponde al rezago espacial que es la interacción del valor (ISAE) en función del peso ponderado de los vecinos incluidos $j \in J_i$.

2. RESULTADOS

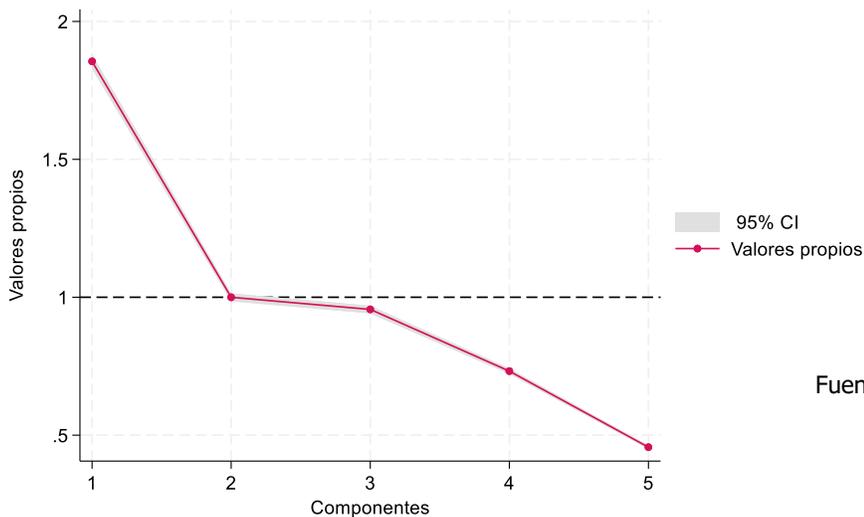
La estimación por componentes principales se realizó utilizando para ello STATA 18, obteniendo los siguientes valores:

Tabla 1. Valores propios y varianza por componente

Componente	Valor propio	Proporción de varianza	Varianza acumulada
1	1,85526	0,3711	0,3711
2	0,999983	0,2000	0,5711
3	0,956007	0,1912	0,7623
4	0,73223	0,1464	0,9087
5	0,456523	0,0913	1,0000

De acuerdo con el criterio de Kaiser, todas las componentes que cuenten con un valor igual o superior a 1 son utilizadas para realizar los análisis. Sin embargo, esta regla general puede ser flexibilizada cuando, como se observa en la tabla 1, los valores propios de los componentes 2 y 3 tienen valores muy próximos a 1. De tal forma que, para efectos del trabajo, se tomarán 3 componentes que en total representan una varianza acumulada de 0,7623. La figura 1 muestra gráficamente los valores propios.

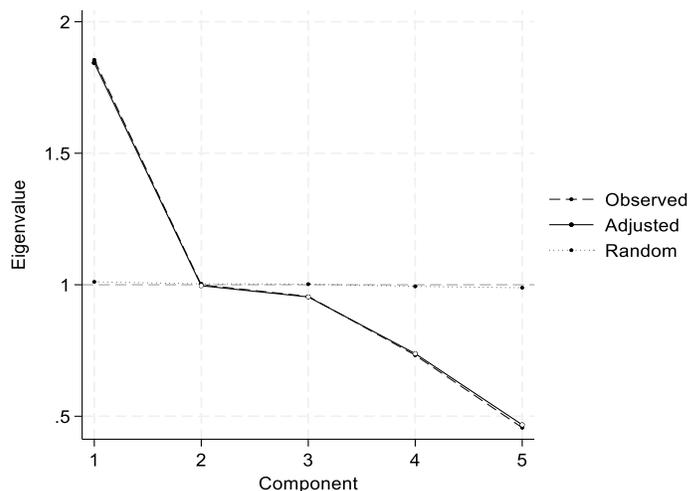
Figura 1. Screeplot de los valores propios por componente



Fuente: cálculos propios

Para validar los resultados obtenidos con el análisis de componente principal respecto a los valores propios, se realiza la estimación de Horn logrando como resultado que el sesgo es muy bajo y por ende la curva observada y ajustada son prácticamente iguales. La figura 2 contiene la gráfica de la iteración.

Figura 2. Iteración de Horn del PCA calculado



Fuente: cálculos propios

Finalmente se realiza la rotación varimax y determinación de los pesos de cada variable y componente. La tabla 2 contiene los resultados obtenidos.

Tabla 2. Pesos por variable y componente

	Componente 1	Componente 2	Componente 3
COM21	0,5613	0,0516	-0,0485
IND21	-0,1075	0,8560	0,0013
SER21	0,4053	0,4802	0,0102
ESTAB21	0,7135	-0,1843	0,0289
EMP21	0,0026	0,0018	0,9988

Fuente: cálculos propios

Las variables de metros cuadrados de comercio y número de establecimientos económicos son las dominantes del componente 1, los metros cuadrados construidos en industria y servicios lo son para el componente 2 mientras que la cantidad de empleo reportada hace lo propio en la componente 3.

Al utilizar la ecuación 2 con los pesos de la tabla 2, ponderado por la varianza explicada de cada componente, se realiza el cálculo del ISAE para cada manzana catastral para el año 2021. Es importante recordar que los pesos obtenidos en la tabla 2 se usan también con las variables estandarizadas del año 2023, de tal forma que sea comparables y generar los análisis intertemporales entre 2021 y 2023.

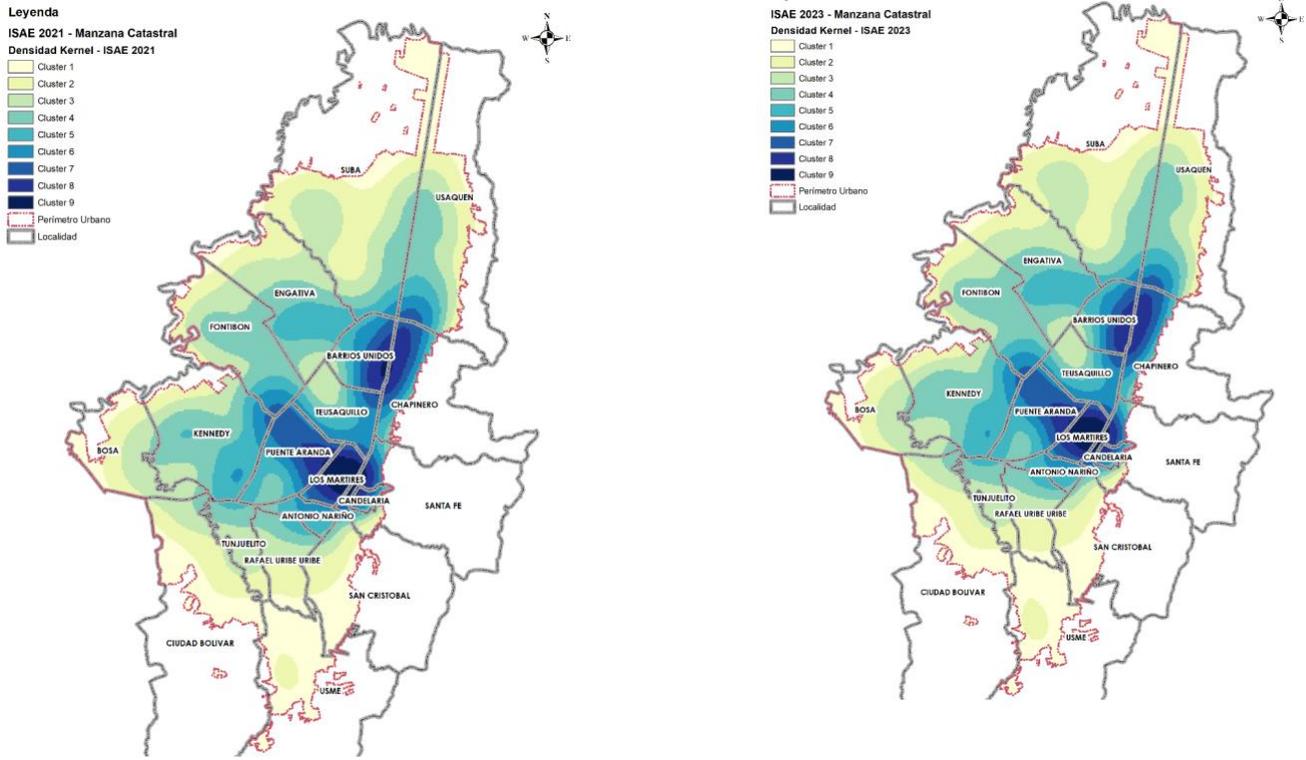
2.1. ANÁLISIS DE DENSIDAD KERNEL INTERTEMPORAL

Ya con los resultados del ISAE a nivel de manzana, se trabajó con la información de los centroides de estos para aplicar el análisis de densidad según los resultados para los años 2021 y 2023. Si bien los mapas cuentan con configuraciones espaciales muy similares (la estructura territorial no va a cambiar de forma abrupta en dos años), si se logran identificar cambio en algunas asociaciones espaciales del ISAE.

Los resultados de la estimación de la densidad kernel a partir del ISAE de manzana censal para los años 2021 y 2023 muestra una pequeña contracción en la actividad económica territorial que se da en las localidades de Kennedy, Engativá, Usaquén y Barrios Unidos. En promedio, la contracción del ISAE es del 1,47%.

Al realizar el análisis por Unidad de Planeamiento Local, unidad administrativa contenida en el Plan de Ordenamiento Territorial, se observa que en promedio la variación positiva se concentra en la zona sur oriental y noroccidental de la ciudad, con la UPL Usme Entrenubes la mayor variación con un 3,4% mientras que las variaciones negativas de mayor valor se dan en las UPL Patio Bonito, Usaquén y Torca.

Figura 3. Análisis de Densidad Kernel ISAE 2021 y 2023



Fuente: elaboración propia

Tabla 3. Variaciones ISAE 2021 – 2023 por UPL

UPL	Variación
Usme - Entrenubes	3,4%
Lucero	3,1%
Cuenca del Tunjuelo	3,0%
Tibabuyes	2,2%
Rafael Uribe	1,9%
Tunjuelito	0,1%
San Cristóbal	0,1%
Tintal	0,0%
Edén	-0,4%
Puente Aranda	-0,7%
Rincón de Suba	-0,8%
Arborizadora	-0,9%
Engativá	-1,1%
Suba	-1,2%
Centro Histórico	-1,3%
Porvenir	-1,9%
Fontibón	-1,9%
Salitre	-2,0%
Niza	-2,2%
Teusaquillo	-2,3%
Tabora	-2,7%
Britalia	-2,8%
Bosa	-2,8%
Chapinero	-3,1%
Restrepo	-4,1%
Cerros Orientales	-4,6%
Toberín	-5,0%
Barrios Unidos	-5,1%
Kennedy	-5,5%

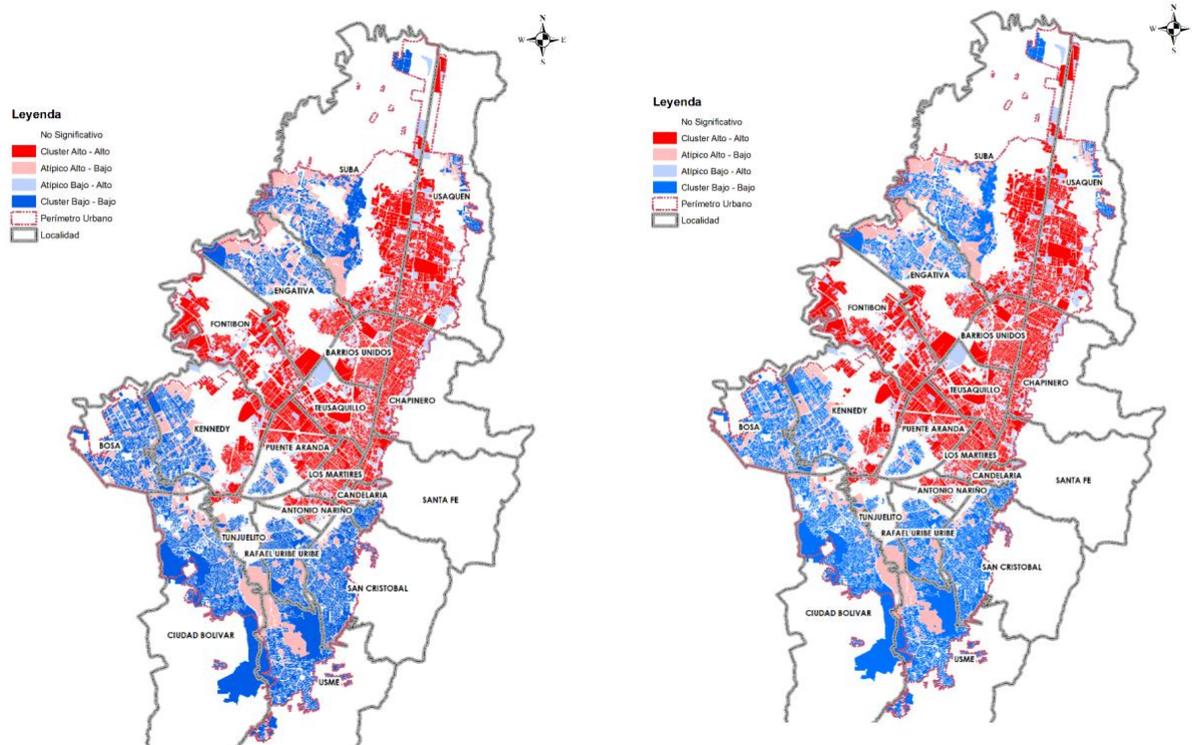
Torca	-6,6%
Usaquén	-8,2%
Patio Bonito	-12,0%

Fuente: elaboración propia

El análisis se complementa con la estimación de los LISA a nivel de manzana catastral, basándose en los resultados del ISAE para cada año. Es importante recordar que la matriz de pesos espaciales que se utiliza es creada con distancia euclidiana y se estandariza por filas.

En general se obtiene un resultado similar al generado con los análisis de densidad, esto es, contar con una estructura territorial muy similar entre los años 2021 y 2023; al analizar el detalle de los *cluster* en las localidades de Teusaquillo, Fontibón y Kennedy. Es importante destacar que, si bien los comportamientos promedio pueden ser positivos o negativo por Localidad o UPL, el análisis LISA estructura aglomeración en función de la cercanía de manzanas catastrales que cuenten con valores altos o el caso contrario.

Figura 4. Análisis LISA del ISAE 2021 y 2023

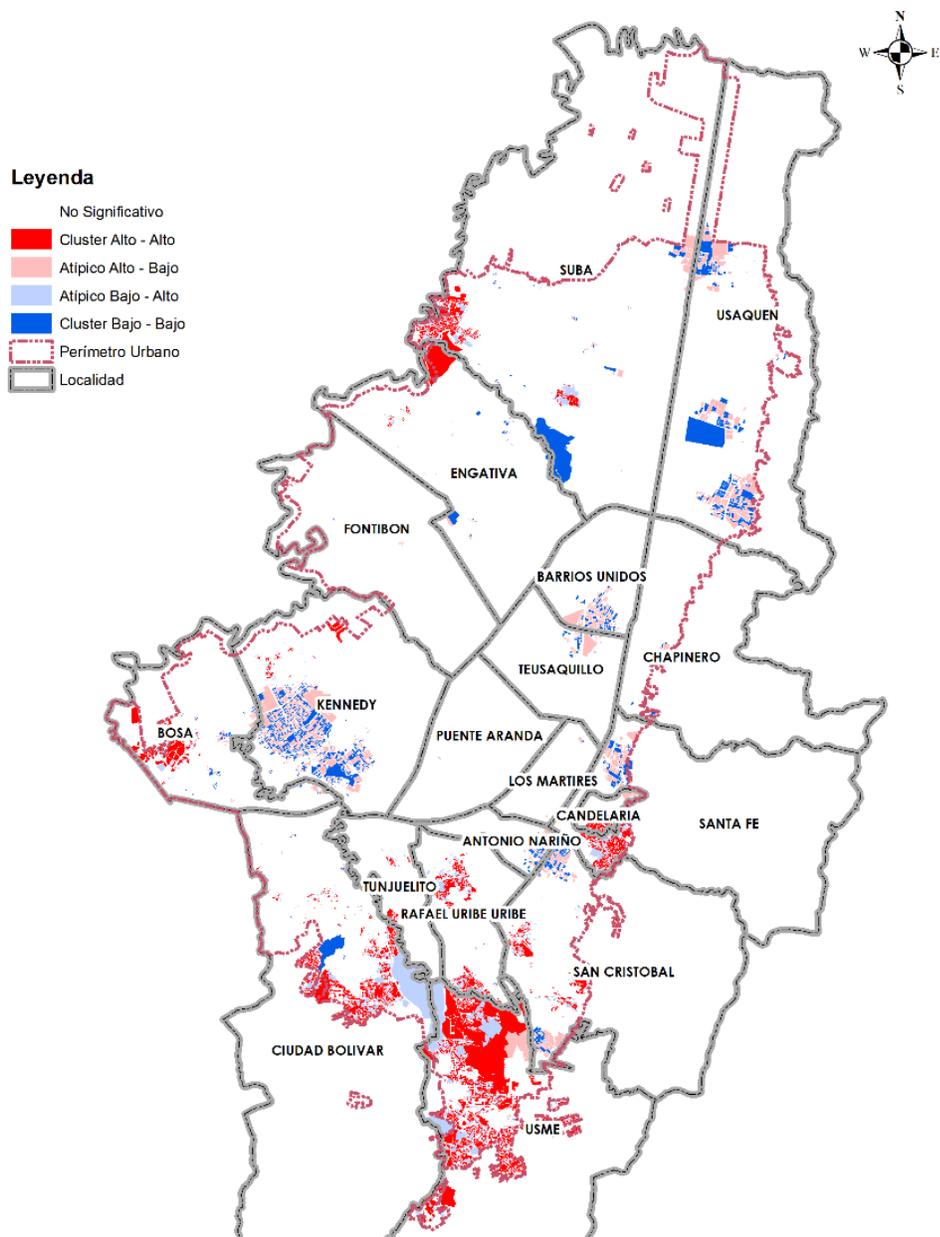


Fuente: elaboración propia

El análisis LISA de las variaciones del ISAE contenido en la figura X permite observar que en las localidades de Bosa, Usme, Ciudad Bolívar, Engativá y Suba se concentran los *cluster* alto alto, lo que representa una concentración de manzanas catastrales de alto valor positivo de variación rodeado de manzanas con la misma condición.

En situación contraria, las localidades de Kennedy, Barrios Unidos, Usaquén y San Cristóbal presentan las mayores concentraciones de *cluster* bajo bajo.

Figura 5. Análisis LISA de la variación del ISAE 2021 y 2023



Fuente: elaboración propia

3. CONCLUSIONES

El ISAE es una medida que busca determinar, a nivel territorial, conocer el comportamiento de la actividad económica. La metodología desarrollada por análisis de componentes principales logra conjugar y dar pesos a cada variable de forma exógena de tal forma que no se equiparen o valoren de forma directa.

El ISAE por si solo es una medida que solo logra jerarquizar las manzanas catastrales objeto de análisis; sin embargo, al dejar los pesos fijos (del año 2021) y volverlo a correr con la estructura de información del año 2023, permite establecer cambios manzana a manzana para conocer la variación intertemporal de la actividad económica.

Los análisis espaciales permiten visualizar en detalle las zonas de la ciudad donde se generaron las mayores y menores variaciones, dirigir acciones de política en polígonos que sean priorizados por la administración según los planes y proyectos contenidos den el POT como puede ser el caso de las Actuaciones Estratégicas.

Sin duda la potencia del ejercicio es la territorialización de las medidas y por ende los análisis que se realizan, que a su vez pueden ser agregado por localidad o UPL. Sin embargo, es necesario lograr desarrollar medidas de comercio informal que puedan complementar las estimaciones; una forma de poder generar esta información es utilizando la información de reconocimiento de imágenes por Google Street View. Es un tema que debe avanzarse a futuro.

4. BIBLIOGRAFÍA

- Anselin, L. (1995). Local Indicators of Spatial Association - LISA. *Geographical Analysis*, 93-115.
- Connor, C., & Connor, L. (2009). Estimating spatial density with kernel methods. En C. Connor, N. Chapman, & L. Connor, *Volcanic and Tectonic Hazard Assessment for Nuclear Facilities* (págs. 346-369). Cambridge: Cambridge University Press.
- Díaz, L., & Morales, M. (2012). *Análisis estadístico de datos multivariados*. Bogotá: Universidad Nacional de Colombia.
- Freudenberg, M. (2003). Composite Indicators of Country Performance: A Critical Assessment. *OECD Science, Technology and Industry Working Papers*, 1-35.
- Silverman, B. (1998). *Density Estimation for Statistics and Data Analysis*. Boca Raton: Chapman & Hall/CRC.
- 