

ESTUDIO DE LA AUTOCORRELACIÓN ESPACIAL EN EL MERCADO INMOBILIARIO COMERCIAL DE LA CIUDAD DE MADRID

STUDY OF SPATIAL AUTOCORRELATION IN THE COMMERCIAL REAL ESTATE MARKET OF THE CITY OF MADRID

Romina Andrea Silvera Vies (Universidad Politécnica de Valencia)^{1*}

Ana Debón (Universidad Politécnica de Valencia)²

Javier Ribal (Universidad Politécnica de Valencia)³

Resumen

La disponibilidad de atributos espaciales hace posible la cuantificación del grado de agrupación o aleatoriedad de una variable en el espacio, ofreciendo así útiles aportes en el análisis exploratorio. En particular, el uso de los índices de autocorrelación espacial permite ahondar en la distribución de una variable tanto a escala global (índice de Moran) como local (LISA), así como la identificación de patrones entre unidades espaciales con nexos de vecindad.

Sin embargo, pese a la vasta utilización de dichos índices en campos como ecología y epidemiología, el empleo de estas técnicas en el análisis de variables en contextos inmobiliarios ha sido limitada, aunque cabe señalar que ayuda a definir la importancia del componente espacial en el valor de un inmueble y si existe una relación clara con los inmuebles del entorno.

En el contexto del mercado inmobiliario nacional, y dada la importancia del comercio minorista en la ciudad de Madrid, uno de los principales núcleos productivos, es de importancia el uso de dichos índices. Por ello, en este trabajo, se ofrece un análisis exploratorio espacial correspondiente a datos de locales comerciales en venta y alquiler en 2020, período caracterizado por el cierre obligatorio del comercio no esencial y limitaciones posteriores en términos aforo, horario de apertura y cierre.

Los resultados muestran patrones entre los 21 distritos que conforman la ciudad, base para el estudio de las rentabilidades brutas de los locales comerciales, que a su vez han sido de los productos inmobiliarios con evolución más favorable en el último año.

Palabras clave: rentabilidad bruta, locales comerciales, autocorrelación espacial, índice de Moran, LISA.

Clasificación JEL: C49, L81

* Autor de correspondencia: romiandreasilvera@gmail.com

¹ ORCID: <https://orcid.org/0009-0005-7589-850X>

² ORCID: <https://orcid.org/0000-0002-5116-289X>

³ ORCID: <https://orcid.org/0000-0002-9355-0145>

Fecha de envío: 30/03/2023. Fecha de aceptación: 29/05/2023

Abstract

The availability of spatial attributes makes it possible to quantify the degree of clustering or randomness of a variable in space, thus offering valuable contributions to exploratory analysis. Particularly, the use of spatial autocorrelation indices makes it possible to delve into the distribution of a variable at both global (Moran's index) and local (LISA) scales and identify patterns between spatial units whose neighborhood links are set according to the hypothesis.

However, despite the extensive use of such indices in fields such as ecology and epidemiology, the use of these techniques in the analysis of variables in real estate contexts has been limited, although it should be noted that it helps to define the importance of the spatial component in the value of a property and whether there is a clear relationship with surrounding properties.

In the context of the national real estate market, and given the importance of retail trade, specifically in Madrid, the use of such indices is essential. Therefore, this paper offers an exploratory spatial analysis corresponding to data on commercial premises for sale and rent in 2020, a period characterized by the compulsory closure of non-essential trade and subsequent limitations in capacity, opening, and closing hours.

The results show patterns among the 21 districts that make up the city, serving as a basis for the study of the gross yields of commercial premises, which in turn have been among the real estate products with the most favorable evolution in the last year.

Keywords: gross yield, commercial premises, spatial autocorrelation, Moran index, LISA.

JEL Codes: C49, L81

1. INTRODUCCIÓN

La existencia de información económica, especialmente en el contexto de bienes inmuebles, hace posible la aplicación de técnicas que permiten evaluar la influencia del componente espacial en la distribución de fenómenos económicos. En este sentido, el uso de modelos espaciales es fundamental para comprender cómo la distribución geográfica de los bienes inmuebles y su entorno pueden afectar a su valor y rendimiento.

La autocorrelación espacial es un concepto estadístico que asigna un valor a la relación entre valores de objetos, dependiendo éstos de vínculos métricos o topológicos. Dicha dependencia, posibilita la estimación de un punto desconocido o la obtención de un modelo estadístico para explicar una variable a partir de otras.

La existencia de este concepto se reflejó en la primera ley de la geografía de Tobler (1970):

«Todo está relacionado a todo lo demás, pero las cosas próximas están más relacionadas que cosas distantes.»

Esta premisa ha sido ampliamente utilizada en campos como la ecología y la epidemiología para cuantificar el grado de agrupación o aleatoriedad de una variable en el espacio (Paez, Farber, y Wheeler, 2011). Los índices de autocorrelación espacial permiten ahondar en la distribución de una variable tanto a escala global como local, así como la identificación de patrones entre unidades espaciales cuyos nexos de vecindad son fijados en función de la hipótesis planteada (Anselin, 1995). En ecología, la autocorrelación espacial se ha utilizado para analizar la distribución de especies en el espacio y la identificación de patrones de diversidad (Legendre, 1993). En epidemiología, se ha utilizado para analizar la distribución espacial de enfermedades y la identificación de áreas de alto riesgo (Waller y Gotway, 2004).

Sin embargo, la autocorrelación espacial también puede ser de gran utilidad en contextos económicos, ya que permite identificar patrones y relaciones espaciales entre objetos y propiedades, lo que puede ser de gran importancia para la toma de decisiones en el mercado inmobiliario, la inversión empresarial y la planificación territorial.

Anselin, Varga y Acs (2002) investigaron los efectos de los vínculos geográficos entre las universidades y las empresas de alta tecnología en la generación de innovaciones en Estados Unidos. Los autores utilizaron técnicas de análisis espacial, incluyendo la autocorrelación, para examinar la relación entre la proximidad geográfica de las empresas y las universidades, y el nivel de innovación tecnológica de las empresas. Los hallazgos sugieren que la proximidad geográfica a una universidad está positivamente relacionada con la capacidad de las empresas para innovar y desarrollar nuevas tecnologías. Este estudio ilustra cómo la autocorrelación espacial puede ser utilizada para identificar patrones y relaciones espaciales en datos económicos y empresariales.

Por otro lado, Arbia y Piras (2015) examinaron la variación espacial de los precios de la vivienda en Bélgica a nivel provincial, también recurriendo a técnicas de análisis espacial, para investigar la presencia de patrones espaciales en los precios de la vivienda y las posibles relaciones con factores socioeconómicos, como la tasa de desempleo, la densidad de población y la calidad del entorno natural. Los autores detectaron que los precios de la vivienda en Bélgica exhiben un fuerte vínculo espacial, lo que indica que la localización geográfica de la vivienda es un factor clave en la determinación de su precio. Además, los resultados sugieren que la tasa de desempleo y la calidad del entorno natural tienen una influencia significativa en los precios de la vivienda. Este estudio destaca la utilidad de la autocorrelación espacial en el análisis de datos económicos y empresariales, así como la importancia de considerar factores espaciales en la comprensión de los patrones y tendencias económicas.

Siguiendo con la línea de aplicación de modelos espaciales en el mercado inmobiliario, se puede mencionar el estudio realizado por McIlhatton et al. (2016), que aportó evidencia del impacto negativo del crimen en el valor de las viviendas, y se encontró que dicho efecto era complejo, variando por tipo de crimen. Por otro lado, se determinó en el estudio de Wu, Guo, y Niu, (2023) que la política de límite de compra de vivienda tenía un efecto significativamente positivo en los precios de estas, específicamente en ciudades grandes y medianas de China.

El objetivo principal de este trabajo es proveer un análisis exploratorio del mercado inmobiliario del área de estudio, mediante la incorporación del componente espacial. Con el cálculo de los índices de Moran local y global se determina si existe interdependencia en los mercados de venta y alquiler, consecutivamente se analiza la significatividad de los estadísticos. Adicionalmente se proporcionan los gráficos de clúster LISA para visualizar de forma directa qué áreas geográficas comparten patrones similares en los precios de compra y alquiler.

En base a los hallazgos, se espera poder ofrecer recomendaciones y estrategias a inversores y empresas del sector inmobiliario en cuanto a la ubicación de los locales comerciales, tales como la detección de zonas con mayor potencial que a su vez indican áreas donde la compra y el alquiler de propiedades pueden tener una mayor rentabilidad.

Por otro lado, conocer qué distritos comparten relaciones de dependencia espacial significativas, en cuanto a los precios de locales comerciales, permite prever comportamientos y tendencias en el mercado inmobiliario, lo que a su vez permite a los inversores y empresas del sector tomar decisiones informadas sobre la ubicación y precio de los locales comerciales en venta o alquiler.

En la siguiente sección, se exponen algunas de las aplicaciones de los métodos utilizados en el análisis, con el fin de contextualizar su uso en el ámbito de estudio. Posteriormente, se detalla la metodología utilizada, y se presenta una explicación de cada uno de los procedimientos

empleados. Luego, se describen los datos utilizados y se presentan los resultados obtenidos a través de su análisis. Finalmente, se presentan las conclusiones y las posibles implicaciones para futuros estudios en el área

2. METODOLOGÍA

El análisis espacial considera dos clases de información: los atributos de rasgos espaciales, como pueden ser las medidas (precipitaciones, temperatura, población) y variables cualitativas como: religión, tipo de suelo, etc. Por otro lado, la localización de cada atributo puede ser descrita, como su posición en un mapa, a través de sistemas de coordenadas o cualquier referencia geográfica. La autocorrelación, se considera una de las, relativamente, pequeñas técnicas que se ocupa tanto de atributos como de la ubicación (Goodchild, 1986). El grado o intensidad de autocorrelación espacial en un patrón depende estrechamente de la escala y del nexo vecinal seleccionado.

Siguiendo la naturaleza de otros tipos de análisis, el análisis espacial también aplica el principio del contraste de hipótesis. La cuestión por resolver se expresa mediante el enfrentamiento de dos hipótesis: una hipótesis nula (H_0) y una hipótesis alternativa (H_a). El principio se basa en la comparación de la situación que se observa en los datos bajo H_0 (suponiendo que es cierta) y los datos obtenidos. Se calcula un índice numérico que sintetiza el problema a resolver (test estadístico) que se utiliza para calcular el p-valor (Souris, 2019) y resolver el contraste anterior.

Cuando los datos muestran evidencia clara en contra de los supuestos de la hipótesis nula, la magnitud del estadístico de prueba se vuelve demasiado grande o pequeño dependiendo de la hipótesis alternativa. Esto hace que el p-valor del test se vuelva lo suficientemente pequeño como para rechazar la hipótesis nula.

En este trabajo empleamos dos índices: el índice de Moran y el índice de Moran local.

Es importante destacar que el uso de los índices de Moran y Moran local no es exclusivo de este trabajo, investigadores los han utilizado en estudios inmobiliarios actuales (Yin, Sun y Bi, 2022; Barreca, Fregonara y Rolando, 2021). Estos índices tienen la ventaja de identificar patrones de agrupación espacial, lo que puede ser útil para identificar áreas con un potencial de inversión interesante y para comprender cómo se distribuyen los precios de las propiedades en una zona determinada. Además, permiten detectar áreas con valores extremadamente altos o bajos en relación con su entorno, lo que es importante para evaluar la competitividad de un área en particular.

2.1. Índice de Moran

La herramienta I de Moran es una medida estadística desarrollada por Moran (1948), que analiza la autocorrelación espacial entre valores de vecinos más cercanos. El índice se define como la media de los productos de los valores normalizados de los pares de puntos, ponderados por un «peso espacial» en función de la contigüidad o distancia entre ambos puntos del par:

$$I_{Moran} = \frac{n}{S_0} \frac{\sum_i^n w_{ij} (x_i - \bar{x})}{\sum_i^n (x_i - \bar{x})^2}$$

Donde:

- n es el número de unidades espaciales
- S_0 es la suma de todos los pesos espaciales (w_{ij})
- w_{ij} es el peso entre la unidad espacial i y la unidad espacial j
- x_i es el valor de la variable de interés para la unidad espacial i
- \bar{x} es la media de la variable de interés para todas las unidades espaciales

En el contexto del estadístico de Moran, p_i y p_j representan los valores de la variable de interés en dos unidades espaciales distintas, donde la relación entre estas unidades está dada por la matriz de pesos espaciales W_{ij} . Esta matriz contiene información acerca de las relaciones espaciales entre las unidades, por ejemplo, si dos unidades son vecinas, están a una determinada distancia, tienen una conexión de transporte, entre otras posibles formas de relacionar unidades espaciales. En general, la elección de la matriz de pesos dependerá de la hipótesis que se quiera probar y de la disponibilidad de información espacial.

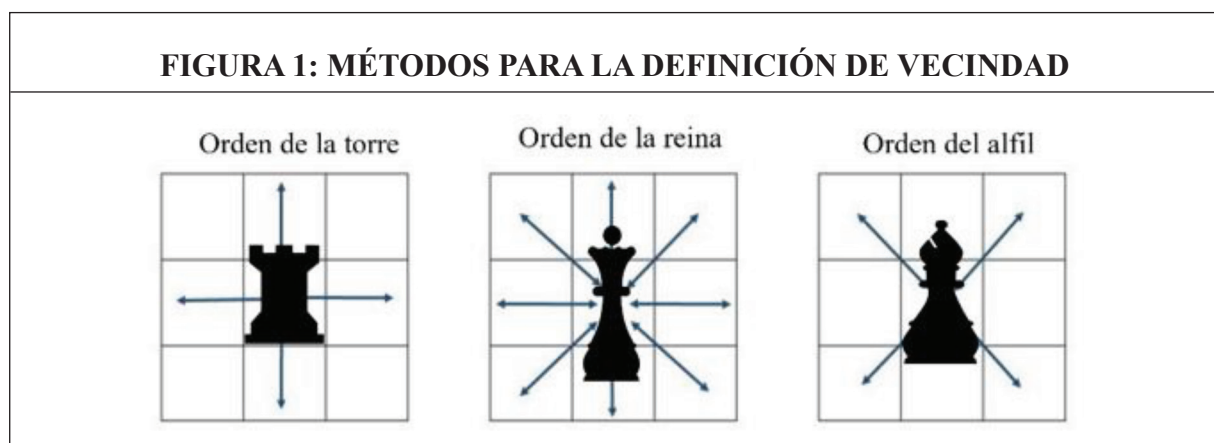
Sus valores oscilan entre -1 y +1, donde +1 significa autocorrelación perfecta (Goodchild, 2009). En definitiva, será positivo cuando áreas cercanas tengan atributos similares y negativo en el caso contrario. En caso de ser nulo, se entiende como la independencia o ausencia absoluta de autocorrelación. Sus resultados siempre son interpretados dentro del contexto de la hipótesis nula. Cuando el p-valor que devuelve esta herramienta es estadísticamente significativo, se puede rechazar la hipótesis nula, que establece que el atributo que se analiza está distribuido de forma aleatoria entre las entidades del área de estudio y por tanto no hay dependencia espacial.

Determinadas las entidades espaciales que se consideran para el análisis de autocorrelación, se define el criterio de vecindad para la creación de la matriz de ponderaciones espaciales W , donde cada entrada w_{ij} es 1 o 0; 1 si el par de regiones son vecinas y 0 en caso contrario. La matriz es de dimensión $n \times n$, donde n es el número de regiones.

Hay otras maneras de definir regiones vecinas, una podría ser la cantidad de fronteras comunes o bien la distancia entre un punto de referencia con cada región (Siabato y Guzmán-Manrique, 2019). Para este estudio se utiliza una relación de vecindad de reina, uno de los criterios de contigüidad más comunes, siendo los principales derivados de movimientos de ciertas piezas de ajedrez como se aprecia en la Figura 1.

Para obtener el estadístico global de todas las observaciones en la muestra, se realiza una prueba analítica para evaluar si hay autocorrelación espacial significativa en un conjunto de datos. El test utiliza como estadístico el índice de Moran y se compara el valor obtenido con una distribución conocida de valores bajo la hipótesis nula de ausencia de autocorrelación espacial (Sánchez, 2012). Si el valor observado está muy alejado de esta distribución, se rechaza la hipótesis nula y se concluye que hay autocorrelación espacial significativa.

Para llevar a cabo lo anterior se recurre a la función *moran.test()* de la librería “spdep” en R. Esta función toma como entrada un objeto de datos espaciales (como un objeto de la clase *SpatialPointsDataFrame*, *SpatialPolygonsDataFrame*, *SpatialPixelsDataFrame*, etc.) y calcula el índice de autocorrelación espacial global (Índice de Moran), así como su p-valor asociado.



Complementariamente, hemos empleado la simulación de Monte Carlo. En el contexto de la autocorrelación espacial, se utiliza para generar distribuciones de valores bajo la hipótesis nula de ausencia de autocorrelación espacial y así poder evaluar la significatividad estadística del índice de autocorrelación obtenido a partir de los datos observados (Dale y Fortin, 2002). Al generar múltiples distribuciones aleatorias, se puede obtener una distribución de valores esperados para el índice de autocorrelación, lo que permite establecer un umbral de significatividad para el valor observado y concluir si existe autocorrelación espacial significativa en los datos (Dubin, 2003).

2.2. Índice de Moran Local

La I de Moran local identifica agrupaciones espaciales o valores atípicos para comprender la contribución de cada unidad espacial al estadístico. Fue desarrollado por Anselin (1995) y se considera un tipo de Indicador Local de Asociación Espacial (LISA). Los LISA tienen dos propósitos en el análisis de datos espaciales exploratorios (ESDA): indican grupos espaciales locales y realizan análisis de sensibilidad (identifican valores atípicos).

Anselin (1995) aporta las siguientes definiciones sobre los indicadores locales:

- El LISA para cada observación da una indicación de la extensión del agrupamiento espacial significativo de valores similares alrededor de cada observación.
- La suma de LISA para todas las observaciones es proporcional a un indicador global de asociación espacial.

Al igual que con las medidas globales, evidencian si el patrón espacial observado de una variable de interés entre áreas es extremo, probable o esperado, dada una distribución geográfica aleatoria de la variable.

Los LISA son técnicas exploratorias basadas en sistemas de información geográfica (GIS) que permiten estudiar la variabilidad espacial de elementos locales, incluso cuando no hay autocorrelación global. Son útiles para grandes áreas con diferencias regionales marcadas. Los LISA miden la concentración de valores numéricos tomados de un conjunto de puntos y se calculan para cada punto usando los vecinos de ese punto. El resultado puede indicar un "hot spot" (punto de alta concentración), un "cold spot" (punto de baja concentración) o un "distinctive spot" (punto rodeado de valores opuestos).

El índice de Moran local (Anselin, 1995) es una descomposición del índice de Moran global. Para cada punto, se puede escribir lo siguiente:

$$I_{Moran(pi)} = \frac{x_i - \bar{x}}{S^2} \sum_{i=1, i \neq j}^n w_{ij} (x_j - \bar{x})$$

Donde:

- x_i es el valor de la variable de interés en la unidad espacial i .
- \bar{x} es la media de los valores de la variable de interés en todas las unidades espaciales de la región.
- S^2 es la varianza de la variable de interés en todas las unidades espaciales de la región.
- w_{ij} es el peso espacial asociado con la relación entre la unidad espacial i y la unidad espacial j .

Esencialmente, este índice permite diferenciar situaciones estadísticamente significativas por simulación, como en el caso general, aunque la simulación se refiere aquí sólo a los puntos implicados en el cálculo local (Wulder y Boots, 1998).

Su interpretación es directa, los valores positivos de I_i sugieren la existencia de un clúster espacial, donde valores similares se agrupan (sean altos o bajos), y los valores negativos de I_i representan un outlier con valores no análogos.

3. DESCRIPCIÓN DE DATOS

El análisis se centra en el mercado inmobiliario de locales comerciales en la ciudad de Madrid, tanto del sector de venta como de alquiler. Los locales comerciales son espacios destinados a la actividad empresarial, y pueden ser utilizados para diferentes fines, como tiendas, oficinas, restaurantes, entre otros.

El mercado de venta de locales comerciales se refiere a las transacciones de compraventa de dichos inmuebles entre particulares o empresas, mientras que el mercado de alquiler se refiere a las transacciones de arrendamiento entre propietarios e inquilinos. Ambos mercados son de gran importancia para el sector inmobiliario, ya que permiten conocer la oferta y la demanda de locales comerciales en una determinada zona.

La fuente de datos consultada fue el portal inmobiliario Fotocasa, de carácter líder en el ámbito nacional y fundado en 1999. La información fue obtenida a través de un proceso de *web scrapping*, que posibilita el acceso a grandes cantidades de datos de distinta estructura para su almacenamiento, posterior tratamiento y aplicación de técnicas de minería (Gandrud, 2020). La descarga de datos se realizó el 20 de febrero de 2021. El tamaño de la muestra obtenida fue de 1.547 locales comerciales en venta y 955 locales en alquiler.

Tras la depuración de los ficheros de venta y alquiler, ambos *dataframes* se han reducido para incluir solo los campos relevantes al estudio:

1. Número identificador (**id**) de cada inmueble en el portal inmobiliario.
2. Los puntos de **latitud** y **longitud** indican la posición de los inmuebles, se hallan en campos individuales y son esenciales para la realización de las técnicas espaciales.
3. Se cuenta con información de naturaleza descriptiva como la **ubicación** y **descripción** de los locales comerciales, ambas expresadas como cadenas de texto y útiles para aplicación de técnicas de *text mining* (Silge y Robinson, 2017).
4. La variable **precio** es una de las principales variables utilizadas en este estudio, ya que permite analizar la relación entre los precios de los locales comerciales y su ubicación geográfica. El tipo de dato es numérico y su sentido varía en función del mercado, en el fichero de venta expresa el importe de adquisición de la propiedad mientras que, en el fichero de alquiler, indica el importe mensual a cambio de la explotación del inmueble.
5. La **superficie** en metros cuadrados también está presente como una de las variables cuantitativas disponibles.

Como ya se mencionó, las coordenadas asociadas con cada inmueble se obtuvieron de Fotocasa, sin embargo, es importante tener en cuenta que las mismas no son exactas, sino que representan una ubicación aproximada dentro de un radio para proteger la privacidad de los propietarios y/o inquilinos. No obstante, se considera que esta imprecisión en la ubicación no tendría un impacto rotundo en los resultados del estudio, ya que solo afectaría a las propiedades que se encuentren justo o muy próximas a la frontera entre dos distritos, y la imprecisión en la ubicación podría asignarlas a uno u otro distrito.

Además de las variables originales obtenidas del portal inmobiliario, se agregará el campo cualitativo “**distrito**” para agrupar los locales por áreas geográficas.

Para llevar a cabo el análisis y visualización de los datos en este estudio, se empleó el lenguaje y entorno computacional R (R. Team, 2022) de carácter estadístico y gráfico. Se define como un proyecto de GNU, sistema operativo de software libre. R pone a disposición del usuario un extensivo rango de técnicas estadísticas y gráficas, dotando de comodidad los procesos de manipulación, cálculo y exhibición de la información. Conjuntamente, se emplea Rstudio (Rs. Team 2021) como entorno de desarrollo integrado (IDE) para R.

Es necesaria la transformación espacial de los campos de latitud y longitud, inicialmente en formato numérico, para poder ser utilizadas en análisis geográficos. Para ello se recurre la librería "sf" de R (Pebesma, 2018), la cual permite trabajar con datos espaciales vectoriales. Con esta librería se crean objetos de tipo "POINT" a partir del par de variables, y se les asigna un sistema de referencia de coordenadas (CRS) en formato EPSG:4326, el cual es un estándar mundial para la representación de datos geográficos en coordenadas geográficas.

De esta forma, se pudo trabajar con estas variables de forma adecuada en los análisis espaciales, permitiendo visualizar los datos en un mapa y realizar cálculos de distancias y áreas.

Para la ejecución del análisis exploratorio y posterior determinación de dependencia geográfica se utiliza un fichero shape del Portal de datos abiertos de la Comunidad de Madrid, que proporciona la división geográfica por distritos de la ciudad.

3.1. Caracterización de la muestra

Mediante el análisis exploratorio y descriptivo de ambos mercados, es posible conocer las características básicas de los datos y realizar comparaciones entre grupos de datos, así como detectar patrones o tendencias (Datar y Harish, 2019).

En las Tablas 1 y 2 se observan distribuciones asimétricas y rangos extremos de precios en ambos mercados, donde el precio de adquisición mínimo es de 12.500 euros y la propiedad más costosa es de casi 9 millones de euros. El importe mensual más bajo para el alquiler es de 190 euros, mientras que la cifra más elevada es de más de 90.000 euros mensuales. Además, ambos conjuntos de datos muestran una elevada dispersión con coeficientes de 150% y 182% para alquiler y venta, respectivamente. La variable superficie también se caracteriza por una marcada asimetría y una gran dispersión, con coeficientes de variación del 215,09% y 324,51%

**TABLA 1: DISTRIBUCIÓN DE LA VARIABLE PRECIO
(EN EUROS) POR MERCADO**

Mercado	Mínimo	Q1	Mediana	Media	Q3	Máximo
Alquiler	190	939	1.592	2.847	2.992	97.196
Venta	12.500	130.000	230.000	434.360	457.000	8.990.000

**TABLA 2: DISTRIBUCIÓN DE LA VARIABLE SUPERFICIE
(EN m^2) POR MERCADO**

Mercado	Mínimo	Q1	Mediana	Media	Q3	Máximo
Alquiler	10	73,8	120,5	266,4	220,0	17.672
Venta	5	73,5	135,0	256,4	254,5	12.500

para alquiler y venta, respectivamente. La información geográfica se incorpora mediante la transformación espacial de las coordenadas y la utilización de la geometría de la ciudad de Madrid, lo que permite visualizar los porcentajes de ofertas por distritos, así como valores globales y unitarios medios (Tabla 3 y Tabla 4).

TABLA 3: CUOTA DE OFERTAS Y PRECIOS MEDIOS POR MERCADO Y DISTRITO

Distrito	Alquiler		Venta	
	Precio medio	% oferta	Precio medio	% oferta
Arganzuela	1.835	5,2 %	481.069	3,6 %
Barajas	2.738	0,9 %	169.080	0,9 %
Carabanchel	2.865	5,5 %	276.894	9 %
Centro	3.584	13,2 %	603.101	7,2 %
Chamartin	2.783	7,3 %	534.747	4,5 %
Chamberi	2.012	9,1 %	500.201	5,9 %
Ciudad Lineal	1.754	5,3 %	342.213	6,7 %
Fuencarral	1.702	4,9 %	591.792	3,6 %
Hortaleza	2.313	3,6 %	410.668	2,6 %
Latina	1.146	4,7 %	263.526	5,4 %
Moncloa	3.094	3,4 %	552.531	3 %
Moratalaz	5.988	1 %	229.388	0,6 %
Puente de Vallecas	982	2,4 %	162.605	6,6 %
Retiro	2.271	3,9 %	591.136	2,8 %
Salamanca	3.318	9,7 %	1.013.921	6,3 %
San Blas	3.078	4,1 %	379.089	5 %
Tetuan	3.235	5,1 %	464.634	8,4 %
Usera	2.191	1,5 %	253.898	3,5 %
Vicalvaro	17.188	0,4 %	166.533	3,6 %
Villa de Vallecas	4.055	5,5 %	345.884	5,6 %
Villaverde	5.678	3,1 %	439.634	5,2 %

TABLA 4: VALORES UNITARIOS MEDIOS (€/m²) POR MERCADO Y DISTRITO

Distrito	Mercado	
	Alquiler	Venta
Arganzuela	14,3	2.388
Barajas	18,7	2.272,6
Carabanchel	9,1	1.263,8
Centro	25,7	4.262,8
Chamartin	16,8	2.859,2

(continuada)

**TABLA 4: VALORES UNITARIOS MEDIOS
(€/m²) POR MERCADO Y DISTRITO (continuada)**

Distrito	Mercado	
	Alquiler	Venta
Chamberí	17,1	3.669
Ciudad Lineal	11	2.011,5
Fuencarral	13	2.276
Hortaleza	11,6	2.333,3
Latina	10,3	1.369,7
Moncloa	18,5	3.271,8
Moratalaz	34,6	2.298,1
Puente de Vallecas	10,3	1.374,1
Retiro	17,9	2.692,9
Salamanca	24,6	4.808,6
San Blas	11,2	1.537,7
Tetuan	18,3	2.263
Usera	11,5	1.401,7
Vicalvaro	6,3	1.828,3
Villa de Vallecas	33,7	1.417,6
Villaverde	9,3	1.205,4

Inicialmente y con base en los sets de datos analizados, se observó que la proporción de locales en venta superaba a la de alquiler en más de un 50%, no obstante, en ciertas zonas marcadamente comerciales de la ciudad, como Centro, Chamberí y Salamanca, la cuota de ofertas es balanceada para ambos mercados, de igual forma se observa un patrón en áreas como Arganzuela, Chamartín, Fuencarral, Moratalaz y Retiro, ya que sus valores unitarios medios se posicionan entre los más elevados de los 21 distritos independientemente del mercado.

En contraste a lo anterior, en zonas con importantes desbalances (predominan las ofertas de venta), los valores unitarios medios se corresponden con los más bajos de la ciudad independientemente si se trate de venta o alquiler, esto se observa en Carabanchel, Ciudad Lineal, Latina, Puente de Vallecas, San Blas y Usera, siendo mayormente áreas residenciales y con menos opciones comerciales en comparación con otras zonas más céntricas de la ciudad. En ningún territorio las ofertas de alquiler superan a las ventas.

Al enfocarse en los precios por metro cuadrado, en el mercado de venta hay notable dispersión (Figura 2), aunque la mayoría de los valores suelen concentrarse entre los 1.000 y 3.000 €/m², presentando una mayor variabilidad se tienen las zonas de Tetuán, Salamanca, Moratalaz, Fuencarral y Chamberí. Algunos distritos como Villaverde, Villa de Vallecas, Usera y San Blas muestran valores unitarios más homogéneos. Cabe destacar que hay valores extremos en todos los distritos, siendo el mayor de 40.000 euros (en Moratalaz).

Los valores unitarios de los locales de alquiler (Figura 3) rondan entre los 8 y 20 euros, siendo el valor más extremo de 230 euros. Tetuán y Salamanca, como en el mercado de venta, reportan una notable amplitud de variabilidad. Observamos que las medianas se posicionan muy cerca del centro en todos los casos excepto en Usera, Moratalaz o Barajas donde la distribución consiste mayoritariamente en precios inferiores a la mediana.

FIGURA 2: PRECIO POR m^2 MERCADO DE VENTA

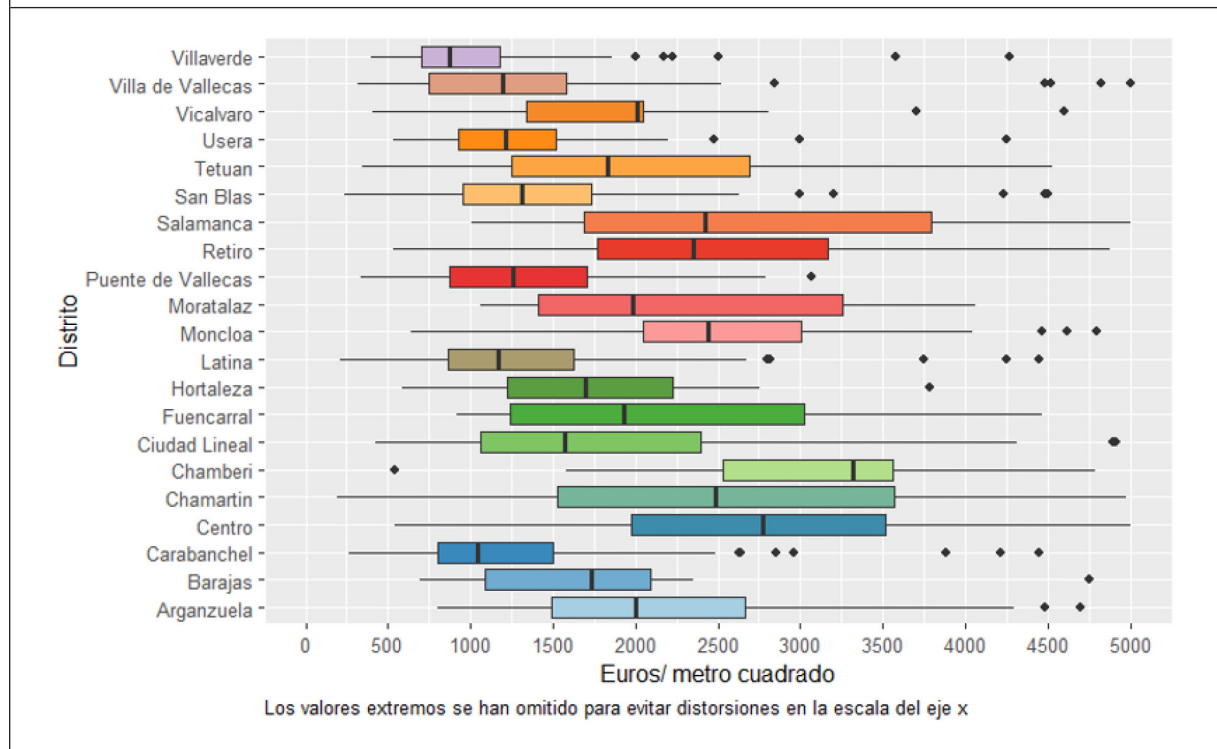
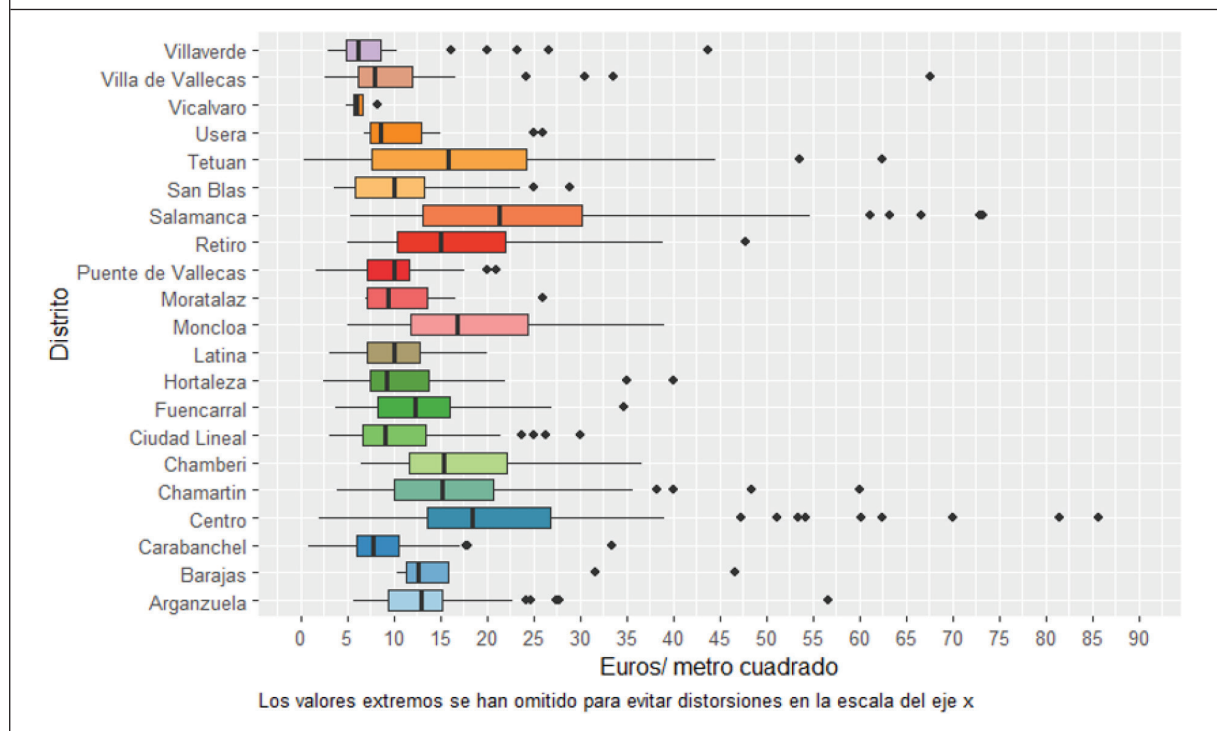


FIGURA 3: PRECIO POR m^2 MERCADO DE ALQUILER



4. APLICACIÓN DE TÉCNICAS

El contraste de hipótesis a realizar sobre ambos mercados puede plantearse como:

H_0 : Los precios de los locales no guardan una relación espacial

H_a : Los precios de los locales comerciales son espacialmente dependientes

Se aplica un método analítico y una simulación Monte Carlo para cada mercado.

Ambos resúmenes analíticos pueden visualizarse en las Tablas 5 y 6, donde los estadísticos para ambos sets de datos son inferiores a la unidad, exactamente 0,00988 en el caso de alquiler y 0,0363 para la venta.

TABLA 5: TEST ANALÍTICO MORAL GLOBAL: MERCADO DE ALQUILER

```

1 globalalquiler
1
2 ## Moran I test under randomisation
3 ## data: alquiler$rawPrice
4 ## weights: lw_alquiler
5
6 ## Moran I statistic standard deviate = 5.785, p-value = 3.626e-09
7 ## alternative hypothesis: greater
8 ## sample estimates:
9 ## Moran I statistic
10
11 ##          Expectation          Variance
11 ##          9.883449e-03    -1.048218e-03    3.570819e-06

```

TABLA 6: TEST ANALÍTICO MORAL GLOBAL: MERCADO DE VENTA

```

1 globalventa
1
2 ## Moran I test under randomisation
3 ## data: venta$rawPrice
4 ## weights: lw_venta
5
6 ## Moran I statistic standard deviate = 26.197, p-value < 2.2e-16
7 ## alternative hypothesis: greater
8 ## sample estimates:
9 ## Moran I statistic
10
11 ##          Expectation          Variance
11 ##          3.692153e-02    -6.468305e-04    2.056528e-06

```

A priori, dichos coeficientes sugieren una inminente independencia espacial. No obstante, antes de descartar la hipótesis alternativa, se debe recurrir a los p-valores obtenidos que indican que, suponiendo la hipótesis nula es cierta, es poco probable que los conjuntos de datos observados se deban al azar ($p\text{-valor} < 0,05$). Si bien ya se ha mencionado la importancia de la escala durante la identificación de patrones, se procede a realizar simulaciones de Montecarlo, cuyo objetivo es corroborar la significatividad estadística del contraste de hipótesis.

Los procesos anteriores se visualizan en las Figuras 4 y 5 mediante curvas representando los valores esperados de los estadísticos de Moran si los precios realmente siguieran un patrón aleatorio. Nótese que ambos Moran I *statistic*, se posicionan a la derecha de la distribución aleatoria, pero la distancia en los datos de venta es mucho mayor a la de los datos de alquiler, indicando una mayor agrupación en los precios de venta.

FIGURA 4: DENSIDAD DE RESULTADOS DE PERMUTACIÓN (MERCADO DE ALQUILER)

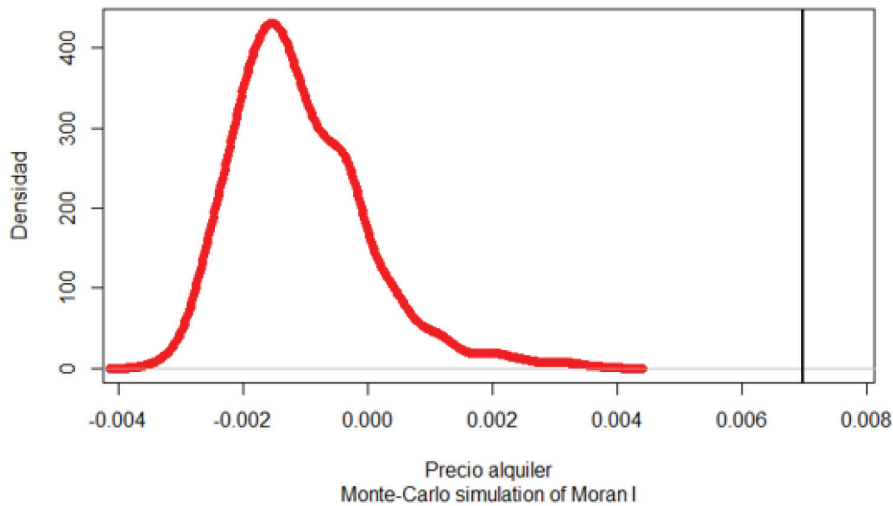
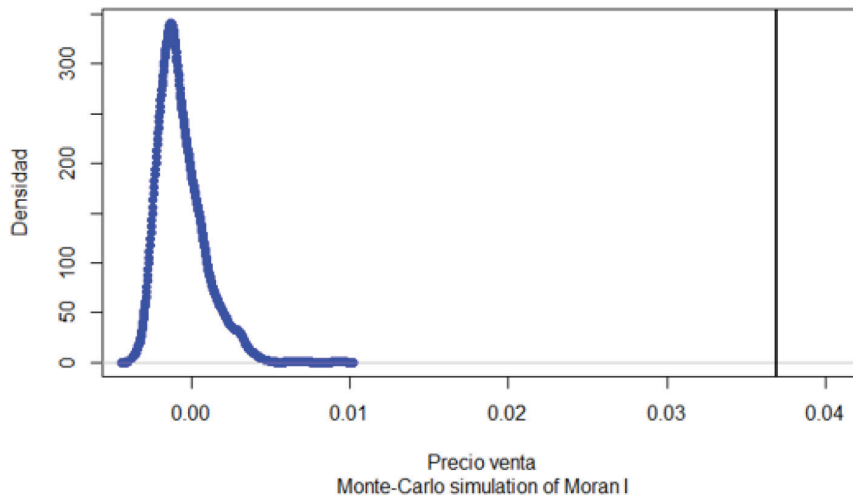


FIGURA 5: DENSIDAD DE RESULTADOS DE PERMUTACIÓN (MERCADO DE VENTA)

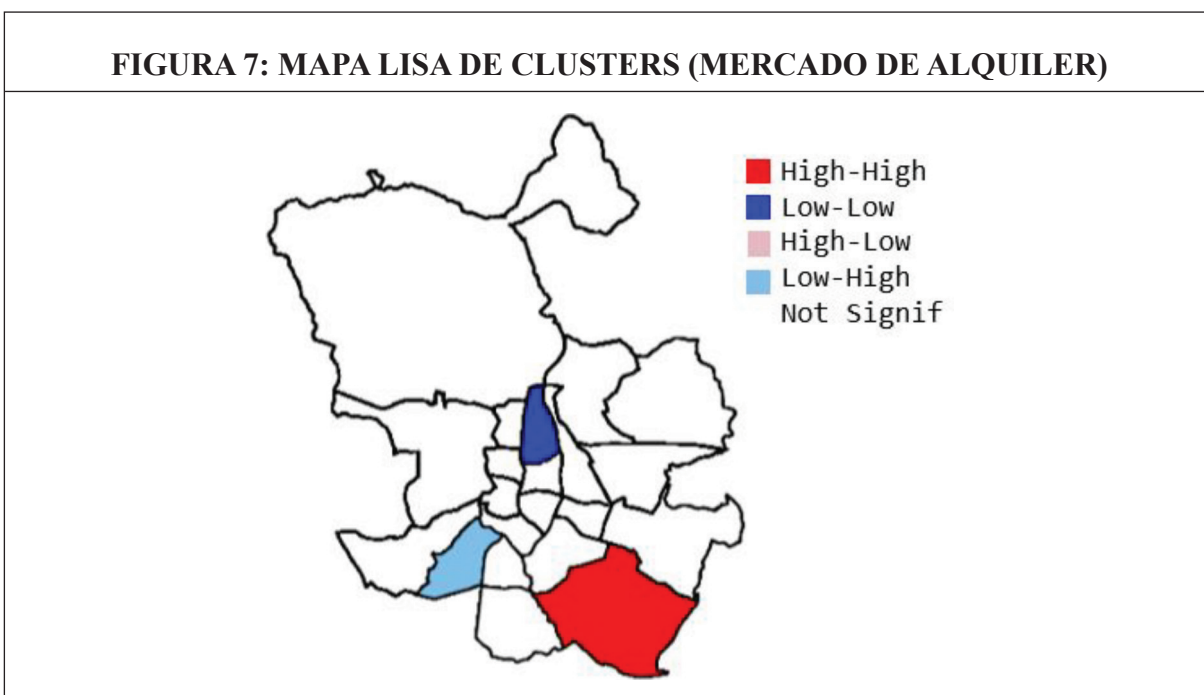
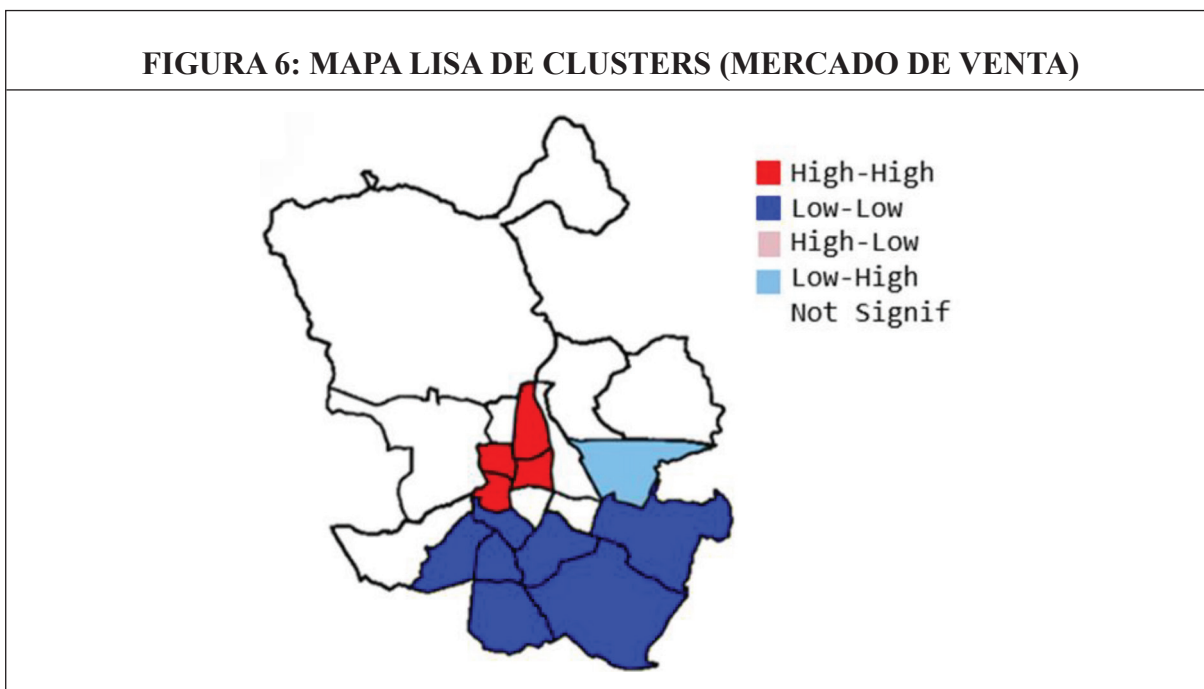


Para confirmar la presencia de asociación a nivel local, se ilustran las Figuras 6 y 7 con los clústeres por unidades espaciales que presentan patrones atípicos respecto a sus vecinos.

En el mercado de venta (Figura 6) se observa la existencia de diversos clústeres espaciales:

Precios superiores a la media global rodeados de precios también siguen dicho comportamiento (*hot spots*), dispuestos en la zona central de la ciudad donde encontramos a Chamberí, Centro, Chamartín y Salamanca.

Precios inferiores a la media global rodeados de similares (*cold spots*) al sur de la ciudad. Dicho clúster comprime a: Carabanchel, Usera, Arganzuela, Villaverde, Villa de Vallecas, Puente de Vallecas y Vicálvaro.



San Blas-Canillejas sigue un patrón atípico en cuanto a los precios de venta de locales comerciales. Presenta un comportamiento Low-High, lo cual traduce a que sus valores son inferiores a la media global y se encuentran rodeados por valores opuestos (que superan a la media).

En cuanto a los demás distritos, que presentan un color blanco, podemos decir que sus precios no siguen ningún patrón uniforme, o al menos, no tan significativo como los anteriormente nombrados.

Para el mercado del alquiler (Figura 7), no vemos patrones uniformes que aglomeren a más de un distrito, sino más bien comportamientos atípicos de carácter individual como el *hot spot* en Villa de Vallecas.

Lo opuesto ocurre en Chamartín, existiendo homogeneidad significativa en los precios, que son bajos.

Por último, Carabanchel representa un punto Low-High, queriendo decir que sus precios presentan disparidad intra distrital.

5. CONCLUSIONES

En primer lugar, este trabajo presenta una metodología para el estudio de la dependencia espacial de locales en venta y alquiler (mercado inmobiliario comercial) que puede ser aplicado a otras áreas geográficas diferentes de la estudiada.

Tras el estudio de datos reales de locales en alquiler y venta de la ciudad de Madrid gracias al empleo de herramientas disponibles en el entorno de programación R, ha sido posible conocer la estructura por tipologías, cuotas de ofertas por distrito, distribución de precios en el territorio y demás características esenciales para poder brindar un análisis de mercado donde algunos puntos a destacar han sido:

- Centro es el distrito con más cantidad de inmuebles comerciales ofertados, con una cuota conjunta de venta y alquiler superior al 20%, siendo mayor la de alquiler con 13 %.
- La oferta de locales en alquiler se concentra fuertemente en el área central de la ciudad (aproximadamente 32%), específicamente, en Centro, Chamberí y Salamanca.
- Carabanchel es el distrito con más locales en venta (9,05%) seguido de Tetuán (8,05%)

En nuestro análisis, los procesos para la obtención de gráficos, tablas y estadísticos son reproducibles y válidos para datos de diferentes periodos. Además, las herramientas de autocorrelación no sólo agrupan observaciones similares, sino que comunican el grado de correlación, por lo que sirven de mucha ayuda a la hora de hallar patrones que no se aprecian a simple vista, este hecho explica su notoria aplicación en campos como la ecología y demografía.

El índice global de Moran, ha resultado en un estadístico próximo a 0, pero significativo por lo que procede el estudio local (LISA) de clústeres. Con el análisis local (LISA) de los precios de venta, se identifican patrones significativos en los precios del año 2020. Los precios más elevados se distribuyeron uniformemente en el centro de la ciudad, más concretamente en Chamberí, Centro, Chamartín y Salamanca. Por otro lado, los precios más bajos se distribuyeron en los distritos del sur.

Respecto al mercado de venta, fue posible la detección de un *outlier* espacial: San Blas-Canillejas, el cual ofrece precios heterogéneos. Mientras que en el mercado de alquiler de locales comerciales, se identifican comportamientos individuales, como el “*hot spot*” en Villa de Vallecas, donde los alquileres seguían una tendencia al alza, un “*cold spot*” en Chamartín, y se determina que Carabanchel es un *outlier* espacial con disparidad en sus precios.

Cabe señalar que, tras contrastar los índices locales, el mercado de alquiler es el que más aleatoriedad presentó en sus importes en comparación al mercado de venta. A este respecto, se recuerda que el período de tiempo en que recaen los datos estudiados fue sumamente excepcional, con marcadas repercusiones no sólo en la oferta comercial de Madrid, también a escala nacional. Los descensos de índices de ventas, sobre todo en el comercio minorista y de la ocupación evolucionaron de forma distinta en los distritos observados, eso explica las diferencias entre los análisis locales de ambos mercados.

Aparte de los resultados y herramientas utilizadas, este estudio tiene limitaciones en relación con el componente temporal, al ser un estudio *cross-section*, cabría esperar un análisis continuado de datos que permitiera seguir validando resultados y monitorear su evolución.

FINANCIACIÓN

Esta investigación no ha recibido financiación externa.

CONTRIBUCIÓN DE LOS AUTORES

"Conceptualización, Ribal, J y Debón, A.; Metodología, Ribal, J y Debón, A.; Obtención de datos, Ribal, J.; Análisis de datos, Silvera, R.; Redacción - Preparación del borrador original, Silvera, R.; Redacción - Revisión y edición, Ribal, J. y Debón, A.; Supervisión, Ribal, J. y Debón A."

REFERENCIAS

- Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research policy*, 31(7), 1069-1085. [https://doi.org/10.1016/S0048-7333\(01\)00184-6](https://doi.org/10.1016/S0048-7333(01)00184-6)
- Anselin, L. (1995). Local Indicators of Spatial Association—LISA. *Geographical analysis* 27 (2), 93-115. <https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>
- Arbia, G., Espa, G., & Giuliani, D. (2015). Analysis of spatial concentration and dispersion. In Handbook of research methods and applications in economic geography. *Edward Elgar Publishing*. <https://doi.org/10.4337/9780857932679.00012>
- Barreca, A., Fregonara, E., & Rolando, D. (2021). EPC Labels and Building Features: Spatial Implications over Housing Prices. *Sustainability*, 13(5), 2838. <https://doi.org/10.3390/su13052838>
- Dale, M. R., & Fortin, M. J. (2002). Spatial autocorrelation and statistical tests in ecology. *Ecoscience*, 9(2), 162-167. <https://doi.org/10.1080/11956860.2002.11682702>
- Datar, R. y Harish G. (2019). Hands-on exploratory data analysis with R: become an expert in exploratory data analysis using R packages. *1st edition. Packt Publishing*.
- Dubin, R. (2003). Robustness of spatial autocorrelation specifications: some Monte Carlo evidence. *Journal of Regional Science*, 43(2), 221-248. <https://doi.org/10.1111/1467-9787.00297>
- Gandrud, C. (2020). *Reproducible Research with R and RStudio*. Chapman & Hall/CRC The R Series. CRC Press. ISBN: 9780429629594. Retrieved from <https://books.google.es/books?id=ezX3DwAAQBAJ>
- Goodchild, M. F. (1986). *Spatial Autocorrelation*. 1st edition. Geo Books, Norwich.
- Goodchild, M. F. (2009). What Problem? Spatial autocorrelation and geographic information science. *Geographical Analysis*, 41(4), 411-417. <https://doi.org/10.1111/j.1538-4632.2009.00769.x>

- Legendre, P. (1993). Spatial Autocorrelation: Trouble or New Paradigm? *Ecology*, 74(6), 1659–1673. <https://doi.org/10.2307/1939924>
- McIlhatton D., McGreal W., Taltavul, P., Adair A., (2016), Impact of crime on spatial analysis of house prices: evidence from a UK city, *International Journal of Housing Markets and Analysis*, Vol. 9 Iss 4 pp. <http://dx.doi.org/10.1108/IJHMA-10-2015-0065>
- Moran, P. A. P. (1948). The interpretation of statistical maps. *J. R. Stat. Soc. Ser. B-statistical Methodol.* 10 (2), 243–251. <https://doi.org/10.1111/j.2517-6161.1948.tb00012.x>
- Páez, A., Farber, S., & Wheeler, D. (2011). A Simulation-Based Study of Geographically Weighted Regression as a Method for Investigating Spatially Varying Relationships. *Environment and Planning A: Economy and Space*, 43(12), 2992–3010. <https://doi.org/10.1068/a44111>
- Pebesma, E. (2018). Simple Features for R: Standardized Support for Spatial Vector Data. *The R Journal*, 10(1), 439–446. <https://doi.org/10.32614/RJ-2018-009>
- Sanchez Navarro, D. (2012). Una Propuesta Metodológica Para La Definición De Mercados Geográficos Relevantes (A Proposed Methodology for Defining Relevant Geographic Markets). Available at SSRN 2221339.
- Siabato, W., & Guzmán-Manrique, J. (2019). La autocorrelación espacial y el desarrollo de la geografía cuantitativa. *Cuadernos De Geografía: Revista Colombiana De Geografía*, 28(1), 1–22. <https://doi.org/10.15446/rcdg.v28n1.76919>
- Silge, J., Robinson, D. (2017). Text Mining with R: A Tidy Approach. *Estados Unidos: O'Reilly Media*. <https://dl.acm.org/doi/10.5555/3165010>
- Souris, M. (2019). “Epidemiology and geography: principles, methods and tools of spatial analysis”. *ISTE*. <https://doi.org/10.1002/9781119528203>
- Team, R. 2022. R: What is R? Vienna, Austria: R Foundation for Statistical Computing. <https://www.r-project.org/about.html/>
- Team, RStudio. 2021. RStudio: Integrated Development Environment for R. Boston, MA: RStudio, Inc. <http://www.rstudio.com/>
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234–240. <https://doi.org/10.2307/143141>
- Waller, L.A., & Gotway, C.A. (2004). Applied Spatial Statistics for Public Health Data. <https://doi.org/10.1002/0471662682>
- Wu, G., Guo, W., & Niu, X. (2023). Spillover effect analysis of home-purchase limit policy on housing prices in large and medium-sized cities: Evidence from China. *PloS one*, 18(1), e0280235. <https://doi.org/10.1371/journal.pone.0280235>
- Wulder, M., & Boots, B. (1998) Local spatial autocorrelation characteristics of remotely sensed imagery assessed with the Getis statistic, *International Journal of Remote Sensing*, 19:11, 2223-2231, <https://doi.org/10.1080/014311698214983>
- Yin, Z., Sun, R., & Bi, Y. (2022). Spatial-Temporal Change Trend Analysis of Second-Hand House Price in Hefei Based on Spatial Network. *Computational intelligence and neuroscience*, 2022, 6848038. <https://doi.org/10.1155/2022/6848038>