



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Dpto. de Estadística e Investigación Operativa
Aplicadas y Calidad

Huella de carbono y sostenibilidad urbana: un análisis
estadístico espacial en los barrios de Valencia

Trabajo Fin de Máster

Máster Universitario en Ingeniería de Análisis de Datos, Mejora de
Procesos y Toma de Decisiones

AUTOR/A: Toral Heredia, Marta

Tutor/a: Carot Sierra, José Miguel

CURSO ACADÉMICO: 2023/2024

RESUMEN

El presente Trabajo Fin de Máster (TFM) tiene como objetivo analizar y comparar la huella de carbono de los distintos barrios de la ciudad de Valencia mediante la aplicación de técnicas univariantes de análisis estadístico espacial. Adicionalmente, se emplean metodologías de análisis multivariante, como el Análisis de Componentes Principales (PCA) y técnicas de agrupación (clustering con K-means, método jerárquico de Ward y regionalización), para identificar posibles relaciones entre las variables y patrones ocultos en los datos. A partir de la definición de una serie de indicadores individuales y agrupados que miden los Gases de Efecto Invernadero (GEI) generados en cada barrio, el análisis de la autocorrelación espacial revela que no existe una relación espacial significativa entre los barrios de la ciudad a nivel agrupado; sin embargo, sí se observan diferencias significativas entre barrios cuando se examinan con más detalle los indicadores individuales. El fundamento teórico de este proyecto se sustenta en la teoría de “ciudad de los 15 minutos”, la cual busca mejorar la sostenibilidad y la calidad de vida reduciendo la dependencia del automóvil y promoviendo un estilo de vida más local y de fácil acceso a los servicios esenciales. Los hallazgos de este estudio proporcionan una herramienta de ayuda para las instituciones competentes, facilitando la toma de decisiones en materia de contaminación y gestión del medioambiente a nivel de barrio.

Palabras clave: Huella de carbono, Gases de Efecto Invernadero, análisis estadístico espacial, ciudad de los 15 minutos, PCA, clustering, regionalización.

ABSTRACT

The objective of this Master's Thesis (TFM) is to analyze and compare the carbon footprint of different neighborhoods in the city of Valencia by applying univariate techniques of spatial statistical analysis. Additionally, multivariate analysis methodologies such as Principal Component Analysis (PCA) and clustering techniques (K-means clustering, Ward's hierarchical method, and regionalization) are employed to identify possible relationships between variables and hidden patterns in the data. Based on the definition of a series of individual and grouped indicators that measure the Greenhouse Gases (GHG) generated in each neighborhood, the analysis of spatial autocorrelation reveals that there is no significant spatial relationship between the neighborhoods of the city at the aggregated level. However, significant differences between neighborhoods are observed when individual indicators are examined in more detail. The theoretical foundation of this project is based on the “15-minute city” theory, which aims to improve sustainability and quality of life by reducing car dependency and promoting a more local lifestyle with easy access to essential services. The findings of this study provide a helpful tool for competent institutions, facilitating decision-making regarding pollution and environmental management at the neighborhood level.

Keywords: Carbon footprint, Greenhouse Gases, spatial statistical analysis, 15-minute city, PCA, clustering, regionalization.

RESUM

El present Treball fi de màster (TFM) té com a objectiu analitzar i comparar la petjada de carboni dels diferents barris de la ciutat de València mitjançant l'aplicació de tècniques univariants d'anàlisi estadística espacial. Addicionalment, s'empren metodologies d'anàlisi multivariant, com l'Anàlisi de Components Principals (PCA) i tècniques d'agrupació (clustering amb K-means, mètode jeràrquic de Ward i regionalització), per a identificar possibles relacions entre les variables i patrons ocults en les dades. A partir de la definició d'una sèrie d'indicadors individuals i agrupats que mesuren els Gasos d'Efecte d'hivernacle (GEI) generats en cada barri, l'anàlisi de l'autocorrelació espacial revela que no existix una relació espacial significativa entre els barris de la ciutat a nivell agrupat; no obstant això, sí que s'observen diferències significatives entre barris quan s'examinen amb més detall els indicadors individuals. El fonament teòric d'este projecte se sustenta en la teoria de “ciutat dels 15 minuts”, la qual busca millorar la sostenibilitat i la qualitat de vida reduint la dependència de l'automòbil i promovent un estil de vida més local i de fàcil accés als servicis essencials. Les troballes d'este estudi proporcionen una ferramenta d'ajuda per a les institucions competents, facilitant la presa de decisions en matèria de contaminació i gestió del medi ambient a nivell de barri.

Paraules clau: Petjada de carboni, Gasos d'Efecte d'hivernacle, anàlisi estadística espacial, ciutat dels 15 minuts, PCA, clustering, regionalització.

ÍNDICE

INTRODUCCIÓN	1
OBJETIVOS	3
CAPÍTULO I: ANTECEDENTES	4
1.1. Concepto de ciudad de los 15 minutos	4
1.2. Caso de estudio: la ciudad de Valencia	7
1.3. Indicadores de huella de carbono	11
CAPÍTULO II: METODOLOGÍA	14
2.1. Descripción de los datos	15
2.1.1. <i>Indicadores de huella de carbono</i>	15
2.1.2. <i>VARIABLES SOCIOECONÓMICAS</i>	17
2.2. Análisis de Componentes Principales	18
2.3. Mínimos Cuadrados Parciales	19
2.4. Análisis estadístico espacial	20
2.4.1. <i>Matriz de pesos espaciales</i>	22
2.4.2. <i>Autocorrelación espacial global</i>	24
2.4.3. <i>Autocorrelación espacial local</i>	28
2.5. Métodos de agrupación	31
2.5.1. <i>Clustering de unidades geográficas</i>	31
2.5.2. <i>Regionalización</i>	34
CAPÍTULO III: RESULTADOS	35
3.1. Indicadores individuales de huella de carbono	36
3.1.1. <i>Análisis estadístico espacial</i>	37
3.1.2. <i>Análisis de Componentes Principales</i>	50
3.2. Indicadores agrupados de huella de carbono	53
3.2.1. <i>Análisis estadístico espacial</i>	54
3.2.2. <i>Métodos de agrupación</i>	57
3.3. Relación entre los indicadores agrupados de huella de carbono y las variables socioeconómicas	63
3.3.1. <i>Diagnóstico y validación de observaciones</i>	64
3.3.2. <i>Validación de variables</i>	67
3.3.3. <i>Interpretación de los resultados</i>	70

CONCLUSIONES	74
BIBLIOGRAFÍA	77
ANEXO I	83
ANEXO II	84

ÍNDICE DE FIGURAS

Figura 0.1.- Infografía Misión Climática València 2030	2
Figura 1.1.- Diagrama de la ciudad de los 15 minutos	6
Figura 1.2.- Valencia, Capital Verde Europea 2024	7
Figura 1.3.- División administrativa de la ciudad de Valencia	10
Figura 1.4.- Fuentes y límites de las emisiones de GEI de la ciudad	14
Figura 2.1.- Patrones de autocorrelación espacial	21
Figura 2.2.- Ejemplos de tipos de contigüidad física de primer orden	23
Figura 2.3.- Gráfico de Moran teórico con autocorrelación positiva	30
Figura 2.4.- Dendograma teórico seleccionando $k = 3$ clústeres	33
Figura 3.1.- Mapas coropléticos de cada indicador individual coloreados por los valores obtenidos en cada barrio	37
Figura 3.2.- Número de vecinos cercanos para cada barrio según la contigüidad de tipo Reina	38
Figura 3.3.- Representación de las conexiones de los 70 barrios de Valencia según la matriz de pesos basada en kernel gaussiano con $k = 10$	39
Figura 3.4.- Mapas de los vecinos cercanos asociados a cada barrio empleando la matriz de pesos basada en kernel gaussiano con $k = 10$	40
Figura 3.5.- Gráficos de Moran para los 24 indicadores individuales	41
Figura 3.6.- Gráfico KDE del estadístico I_i de Anselin para cada uno de los indicadores individuales	44
Figura 3.7.- Clasificación de los clústeres de la I_i de Anselin para los indicadores individuales estadísticamente significativos	45
Figura 3.8.- Mapas coropléticos de las variables estadísticamente significativas según la C de Geary	46
Figura 3.9.- Scree Plot del PCA para los indicadores individuales	50
Figura 3.10.- Gráficos de loadings para las tres primeras componentes del PCA	52
Figura 3.11.- Gráficos de scores para las tres primeras componentes del PCA	53

Figura 3.12.- Gráfico de Moran para los 5 indicadores agrupados	54
Figura 3.13.- Gráfico KDE del estadístico I_i de Anselin para los indicadores agrupados	56
Figura 3.14.- Clasificación de los clústeres de la I_i de Anselin para los indicadores agrupados.....	56
Figura 3.15.- Mapas coropléticos de los indicadores agrupados.....	56
Figura 3.16.- Comparación coeficiente de Silhouette para clustering K-Means y Ward	58
Figura 3.17.- Clústeres obtenidos según el metodo K-Means (izquierda) y Ward (derecha)	59
Figura 3.18.- Coeficiente de Silhouette para el análisis de regionalización.....	60
Figura 3.19.- Clústeres obtenidos del análisis de regionalización.....	60
Figura 3.20.- R^2 y Q^2 acumuladas para cada componente principal del modelo PLS ..	64
Figura 3.21.- Gráfico SPE del modelo PLS con 6 componentes principales	65
Figura 3.22.- Gráfico de contribuciones para el barrio de La Seu en el modelo PLS...	66
Figura 3.23.- Gráfico de dispersión entre las variables “vehículos autobús” y “turismos particulares”	66
Figura 3.24.- Gráfico T^2 de Hotelling del modelo PLS con 6 componentes principales	67
Figura 3.25.- Coeficientes de regresión para las variables socioeconómicas del modelo PLS	68
Figura 3.26.- Coeficientes de regresión para las variables mantenidas del modelo PLS	69
Figura 3.27.- Score Plot entre los scores de X e Y para las dos componentes principales	70
Figura 3.28.- R^2 de cada variable Y del modelo PLS para cada componente principal	70
Figura 3.29.- R^2 de cada variable X del modelo PLS para cada componente principal	71
Figura 3.30.- Weighting Plot (W^* , $c_1 - W^*$, c_2) del modelo PLS.....	72
Figura 3.31.- Score Plot ($T_1 - T_2$) del modelo PLS coloreado por la huella de carbono total	72
Figura 3.32.- Gráficos para las variables X (izquierda) e Y (derecha) en la zona con valores elevados en la primera componente	73

ÍNDICE DE TABLAS

Tabla 1.1.- Descripción de los 19 distritos y 87 barrios de Valencia	8
Tabla 1.2.- Sectores y subsectores de las emisiones de GEI de una ciudad	13
Tabla 2.1.- Descripción de los indicadores individuales de huella de carbono	16
Tabla 2.2.- Patrones espaciales y valores teóricos de los estadísticos de autocorrelación espacial global	28
Tabla 3.1.- Resumen estadístico de los indicadores individuales.....	36
Tabla 3.2.- Estadísticos I de Moran y C de Geary para los indicadores individuales ...	42
Tabla 3.3.- Loadings de las seis primeras componentes principales del PCA para los indicadores individuales	51
Tabla 3.4.- Estadísticos I de Moran y C de Geary para los indicadores agrupados	55
Tabla 3.5.- Valores de AMI de similitud entre los tres métodos de análisis clúster	61
Tabla 3.6.- Valores de IPQ de coherencia geográfica para los clústeres obtenidos con los tres métodos de clustering.....	62
Tabla 3.7.- Valores del score de CH de bondad de ajuste para los tres métodos de clustering	62
Tabla 3.8.- Resumen R^2 y Q^2 explicado por cada componente principal extraída del modelo PLS definitivo.....	69

INTRODUCCIÓN

El 25 de septiembre de 2015 se ratificó el mayor plan de acción global destinado a “lograr un futuro mejor y más sostenible para todos” (United Nations, 2015). Este acuerdo, denominado Agenda 2030, fue firmado por los 193 países miembros de la Organización de las Naciones Unidas (ONU).

La Agenda 2030 nace como respuesta a la necesidad de abordar los desafíos mundiales más críticos, como son la pobreza, la desigualdad, el cambio climático o el desarrollo sostenible. Esta hoja de ruta plantea 17 Objetivos de Desarrollo Sostenible (ODS) con 169 metas interrelacionadas e indivisibles que abarcan las dimensiones económica, social y medioambiental (United Nations, 2015) y que deberán ser alcanzadas para el año 2030. Para la consecución de estos objetivos, es fundamental que todos los países, instituciones, empresas y la sociedad en general colaboren conjuntamente.

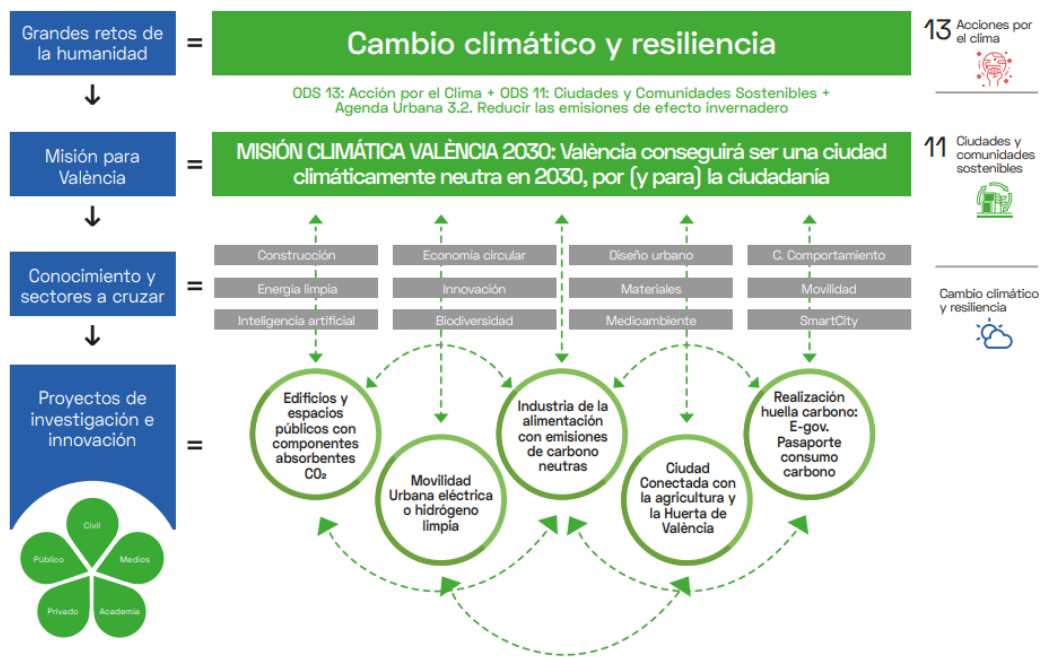
En este contexto, la ciudad de Valencia decidió involucrarse de lleno en estos aspectos lanzando en el año 2020 *Missions València 2030*, un modelo de gobernanza e innovación basado en misiones que mejoren la vida de los ciudadanos de la ciudad (*Missions València 2030*, 2024). Su primera misión es la *Misión Climática València 2030*, cuyo propósito es conseguir que la ciudad sea climáticamente neutra en 2030, en el contexto de la misión europea de “100 ciudades europeas inteligentes y climáticamente neutras en 2030, por y para los ciudadanos” (European Commission et al., 2020).

Basado en esta misión, Valencia aspira a llevar a cabo una reforma estructural en los barrios de la ciudad que les permita absorber el 100% de las emisiones de CO₂ generadas en dichas regiones antes del año 2030. Con este propósito, la ciudad de Valencia contribuye de manera directa en los objetivos y metas establecidos en la Agenda 2030, especialmente en los siguientes:

- ODS 11: “Ciudades y comunidades sostenibles”.
- ODS 13: “Acción por el clima”.

La Figura 0.1 muestra el camino a seguir por la ciudad, cuyo modelo está basado en las ideas de la economista Mariana Mazzucato (2018).

Figura 0.1.- Infografía Misión Climática València 2030



Fuente: Missions València 2030 (2024)

En línea con esta iniciativa, el presente proyecto busca respaldar a las entidades responsables y ofrecer un panorama del estado medioambiental actual de la ciudad de Valencia, con la novedad de proporcionar una perspectiva ampliada al explorar las relaciones espaciales entre los diversos barrios, y con el fin último de promover la implementación de medidas comunes más eficaces.

La combinación de técnicas estadísticas con información georreferenciada permite estudiar relaciones espaciales entre observaciones y facilita la identificación de patrones y tendencias que no serían evidentes en un análisis estadístico tradicional. En los últimos años, el empleo de este tipo de técnicas ha experimentado un notable crecimiento debido a la creciente disponibilidad de este tipo de datos espaciales (Gómez Rubio, 2022). Estos avances abren nuevas posibilidades para la toma de decisiones informadas en numerosas disciplinas, como la salud (Keshtkar et al., 2022), la educación (Ghodousi et al., 2020), la economía (Melecky, 2015) y la planificación urbana.

El estudio se ha realizado en el marco del proyecto que se está desarrollando en la Universidad Politécnica de València (UPV), denominado “NEARCITY. Valencia, la ciudad de los 15 minutos”. Si bien esta investigación trasciende este concepto y la

accesibilidad en sí misma, se han generado sinergias y se ha realizado en estrecha colaboración con dicho proyecto.

Este TFM está estructurado como sigue. En el primer capítulo se realiza una conceptualización teórica fundamentada en una revisión de fuentes de información secundarias que aborda la comprensión de los conceptos y las ubicaciones que se emplearán a lo largo del estudio. El segundo capítulo está enfocado en describir los datos y la metodología, lo cual sienta las bases para que, en el tercer capítulo, se expongan y detallen los resultados obtenidos de los análisis implementados. Finalmente, se presentan las conclusiones alcanzadas, las limitaciones del proyecto y posibles futuras líneas de investigación.

OBJETIVOS

El principal objetivo de este Trabajo Fin de Máster es analizar la huella de carbono de los distintos barrios que conforman la ciudad de Valencia y comparar sus analogías y diferencias mediante la aplicación de técnicas multivariantes y de análisis estadístico espacial. A partir de este objetivo principal, se establecen los siguientes objetivos específicos que guiarán el desarrollo integral del estudio:

- Estudiar y observar las diferencias entre el uso de indicadores individuales y agrupados.
- Evaluar la eficiencia de las técnicas univariantes de análisis estadístico espacial en comparación con otras técnicas de análisis multivariante.
- Elaborar un análisis enfocado a la toma de decisiones para las entidades responsables y la ciudadanía.
- Proporcionar una herramienta para la visualización y comunicación de los resultados del análisis espacial.

CAPÍTULO I: ANTECEDENTES

Para poner en contexto y asentar las bases del presente Trabajo Fin de Máster, a lo largo de este capítulo se introducen algunos conceptos clave que es necesario comprender, como el de “ciudad de los 15 minutos”. Asimismo, se describen los principales rasgos que caracterizan a la ciudad objeto de estudio, Valencia. Finalmente, se explican los términos e indicadores de huella de carbono, los cuales serán analizados a lo largo de todo el proyecto.

1.1. Concepto de ciudad de los 15 minutos

Este estudio se enmarca dentro del concepto de “ciudad de los 15 minutos”. El término fue definido por Carlos Moreno en el año 2021, un científico urbano y profesor asociado en la Universidad de la Sorbona en París. En su artículo *Introducing the “15-Minute City”: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities* (Moreno et al., 2021), Moreno propone un modelo de ciudad en la que los residentes de la misma deberían poder acceder a todos los servicios básicos en un trayecto de 15 minutos a pie o en bicicleta desde sus hogares. Esta propuesta se enfoca en la creación de comunidades más sostenibles, habitables y conectadas, reduciendo la dependencia del automóvil y fomentando un estilo de vida más local y equilibrado.

La idea de ciudad de los 15 minutos está inspirada en el concepto de Smart City. Este innovador modelo de planificación urbana está fundamentado en la implementación de diversas tecnologías para optimizar los tejidos urbanos. Mediante el empleo de tecnologías como la Inteligencia Artificial (IA), el Big Data, el Machine Learning o el Internet de las Cosas (IoT), las Smart Cities representan un nuevo marco para el desarrollo urbano que facilita la gestión eficiente de los recursos y mejora la calidad de vida de los ciudadanos (Yan et al., 2023).

Todo ello, a su vez, puede englobarse en el concepto de “cronourbanismo” (Moreno et al., 2021), que describe que la calidad de la vida urbana es inversamente proporcional a la cantidad de tiempo invertido en el transporte, más aún mediante el uso de automóviles. En una ciudad definida como tal, los residentes podrán disfrutar de una mayor calidad de vida, permitiéndoles cumplir eficientemente seis funciones sociales urbanas esenciales para sostener una vida urbana digna. Estos seis servicios básicos son:

1. **Vivienda:** siendo esta dimensión el núcleo del concepto de la ciudad de los 15 minutos, se plantea que los residentes de un barrio deberían tener acceso a los recursos necesarios en un radio de 15 minutos a pie o en bicicleta desde sus hogares.
2. **Trabajo:** los ciudadanos deberían llegar a sus lugares de trabajo en 15 minutos, un objetivo que resulta claramente complejo y difícil de alcanzar en la mayoría de las ocasiones.
3. **Comercio:** abarca todos los establecimientos abiertos que ofrecen los productos básicos necesarios para los ciudadanos, como alimentos, ropa o productos del hogar.
4. **Atención sanitaria:** esta dimensión recopila el acceso de la población a hospitales y centros de salud que ofrecen servicios médicos y atención sanitaria.
5. **Educación:** esta categoría incluye todas las modalidades de educación para los diferentes grupos de edad, tanto educación infantil como educación primaria, secundaria y bachillerato.
6. **Entretenimiento:** recoge todos los servicios y recursos culturales, deportivos y de ocio que la ciudad ofrece a sus ciudadanos.

Para garantizar que estas seis funciones sociales estén cubiertas, es necesario reestructurar el paisaje urbano construido para que integre los cuatro componentes esenciales que se describen a continuación (Moreno et al., 2021).

- **Densidad:** este factor, medido en términos de personas por kilómetro cuadrado, es crucial para el entorno construido en una ciudad. Al planificar una ciudad sostenible en términos económicos, sociales y medioambientales, resulta fundamental determinar el número óptimo de personas que un área determinada puede sostener cómodamente en términos de prestación de servicios urbanos y consumo de recursos.
- **Proximidad:** esta dimensión puede ser entendida tanto desde una perspectiva temporal como espacial y evalúa si los residentes de un barrio determinado pueden acceder fácilmente a las funciones sociales básicas dentro de los nodos radiales de 15 minutos. Está relacionada directamente con la filosofía del "cronourbanismo"

y es crítica para reducir la cantidad de tiempo perdido en los desplazamientos, así como para minimizar los impactos medioambientales y económicos ocasionados por las diferentes actividades desarrolladas.

- **Diversidad:** entendida desde dos perspectivas. Por un lado, la necesidad de disponer, dentro de un mismo barrio, de zonas residenciales, comerciales y de entretenimiento. Por otro lado, el establecimiento de una diversidad cultural, de edades y de clases sociales que faciliten la cohesión e interacción social.
- **Digitalización:** para asegurar la implementación de los otros tres componentes, esta dimensión juega un papel fundamental y está estrechamente vinculada al concepto de Smart City. Dentro de este marco, se deben fomentar las plataformas digitales y el despliegue efectivo de tecnologías que impulsen la participación ciudadana, la inclusividad y la prestación de servicios en tiempo real.

A modo de resumen, en la Figura 1.1; **Error! No se encuentra el origen de la referencia.** se muestran las seis funciones sociales y los cuatro componentes esenciales que se engloban en el concepto de ciudad de los 15 minutos.

Figura 1.1.- Diagrama de la ciudad de los 15 minutos



Fuente: Elaboración propia a partir de *Moreno et al. (2021)*

1.2. Caso de estudio: la ciudad de Valencia

La ciudad de Valencia se ha convertido en un referente de la sostenibilidad, habiendo sido galardonada con el premio *Capital Verde Europea 2024* (European Commission, 2024). Este premio, instituido y gestionado por la Comisión Europea (véase Figura 1.2), se creó con el fin de reconocer a las ciudades de la Unión Europea que destacan e innovan en la gestión del medioambiente y en la mejora del entorno vital de sus ciudadanos. La ciudad que ostenta la capitalidad cada año actúa como modelo de sostenibilidad y comparte sus prácticas ambientales con otras ciudades.

Concretamente, Valencia ha obtenido este título por sus logros en el campo del turismo sostenible, la neutralidad climática y la transición verde justa e inclusiva. El 97% de los habitantes de la ciudad vive a menos de 300 metros de zonas verdes urbanas. Asimismo, la ciudad ha demostrado su compromiso con la mejora de la calidad del aire y la restauración de los ecosistemas naturales, como los ecosistemas dunares y de humedales de *El Racó de l'Olla* (Sorli Guerola, 2023), una zona geológicamente deprimida situada en pleno corazón del Parque Natural de L'Albufera. Destaca también su impulso hacia una producción más saludable, sostenible e inclusiva, gracias a su “Programa Barrio y Alimentación”.

Figura 1.2.- Valencia, Capital Verde Europea 2024



Fuente: *Universitat Politècnica de València* (2022)

Según el último censo de población publicado por el Instituto Nacional de Estadística con fecha 31 de diciembre de 2023, Valencia tiene 807.693 habitantes (INE, 2023). Es la tercera ciudad más poblada de España, solo superada por Madrid y Barcelona, con

3.332.035 y 1.660.122 habitantes, respectivamente. El término municipal de Valencia tiene una superficie de 134,65 km² y una densidad de población de 5.998 habitantes/km².

Administrativamente, la ciudad de Valencia se divide en 19 distritos, que a su vez se encuentran subdivididos en 87 barrios y pedanías. La Tabla 1.1 recoge la información acerca de estas divisiones administrativas, detallando el número de habitantes y su superficie en hectáreas (ha).

Tabla 1.1.- Descripción de los 19 distritos y 87 barrios de Valencia

Distrito	Barrio	Población	Superficie (ha)
1. Ciutat Vella	1.1. La Seu	2731	22,1
	1.2. La Xerea	3884	31,8
	1.3. El Carme	6017	38,4
	1.4. El Pilar	3986	16,2
	1.5. El Mercat	3263	17,3
	1.6. Sant Francesc	5665	43,9
2. Eixample	2.1. Russafa	25599	87,8
	2.2. El Pla del Remei	7107	38,7
	2.3. Gran Vía	12425	46,9
3. Extramus	3.1. El Botànic	6526	37
	3.2. La Roqueta	4768	23,1
	3.3. La Petxina	15554	49,7
	3.4. Arrancapins	23838	87,4
4. Campanar	4.1. Campanar	12448	98,8
	4.2. Les Tendetes	5715	25,8
	4.3. El Calvari	5270	9,4
	4.4. Sant Pau	11275	389,8
5. La Saïdia	5.1. Marxalenes	11648	39
	5.2. Morvedre	10651	42,8
	5.3. Trinitat	8410	59,6
	5.4. Tormos	9135	28
	5.5. Sant Antoni	10347	25
6. El Pla del Real	6.1. Exposició	6984	38,2
	6.2. Mestalla	15014	84,4
	6.3. Jaume Roig	6951	19,1
	6.4. Ciutat Universitària	2667	27,7
7. L'Olivereta	7.1. Nou Moles	27234	74
	7.2. Soternes	5212	22,5
	7.3. Tres Forques	9303	52,7
	7.4. La Fontsanta	3655	33,2
	7.5. La Llum	5177	18,7
8. Patraix	8.1. Patraix	26032	62,1
	8.2. Sant Isidre	10127	49,5
	8.3. Vara de Quart	11176	105,3
	8.4. Safranar	8691	44,1
	8.5. Favara	3415	28
9. Jesús	9.1. La Raiosa	15789	37,5
	9.2. L'Hort de Senabre	17740	42,6
	9.3. La Creu Coberta	6505	37,5
	9.4. Sant Marcel·lí	10326	33
	9.5. Camí Real	3459	147,8

Tabla 1.1 (cont.). Descripción de los 19 distritos y 87 barrios de Valencia

10. Quatre Carreres	10.1.	Mont-Olivet	20135	47,3
	10.2.	En Corts	12335	35,4
	10.3.	Malilla	22975	250,7
	10.4.	Fonteta Sant Lluís	3073	23,9
	10.5.	Na Rovella	9050	54
	10.6.	La Punta	2183	621,4
	10.7.	Ciutat de les Arts i les Ciències	5523	98,9
11. Poblat Marítims	11.1.	El Grau	9513	71,7
	11.2.	El Cabanyal-el Canyamelar	20769	134,9
	11.3.	La Malva-rosa	14216	72,9
	11.4.	Beteró	8516	25,5
	11.5.	Natzaret	6475	91,7
12. Camins al Grau	12.1.	Aiora	26049	65,1
	12.2.	Albors	9289	25,6
	12.3.	La Creu del Grau	15137	39,2
	12.4.	Camí Fondo	4644	15,7
	12.5.	Penya-roja	8253	91,1
13. Algirós	13.1.	L'Illa Perduda	9718	23,2
	13.2.	Ciutat Jardí	13850	36,4
	13.3.	L'Amistat	8151	20,7
	13.4.	La Bega Baixa	6352	19,4
	13.5.	La Carrasca	3710	198,1
14. Benimaclet	14.1.	Benimaclet	25171	74,3
	14.2.	Camí de Vera	5891	90
15. Rascanya	15.1.	Orriols	17835	39,6
	15.2.	Torrefiel	26289	70
	15.3.	Sant Llorenç	7736	153,2
16. Benicalap	16.1.	Benicalap	36055	171,9
	16.2.	Ciutat Fallera	6552	49,7
17. Pobles del Nord	17.1.	Benifaraig	945	142,3
	17.2.	Poble Nou	1049	29,2
	17.3.	Carpesa	1254	226,4
	17.4.	Cases de Bàrcena	396	127,7
	17.5.	Mauella	59	106,9
	17.6.	Massarajos	1628	280,1
	17.7.	Borbotó	773	141,8
18. Pobles de l'Oeste	18.1.	Benimàmet	13074	147,7
	18.2.	Beniferri	767	53,4
19. Pobles del Sud	19.1.	El Forn d'Alcedo	1362	120,1
	19.2.	El Castellar-l'Oliverar	7027	299
	19.3.	Pinedo		
	19.4.	El Saler	2514	652,6
	19.5.	El Palmar	1527	824,8
	19.6.	El Perellonet	782	321,2
	19.7.	La Torre	1605	492,8
	19.8.	Faitanar	5067	42,5
		406	490,3	

Fuente: Ayuntamiento de Valencia (2022)

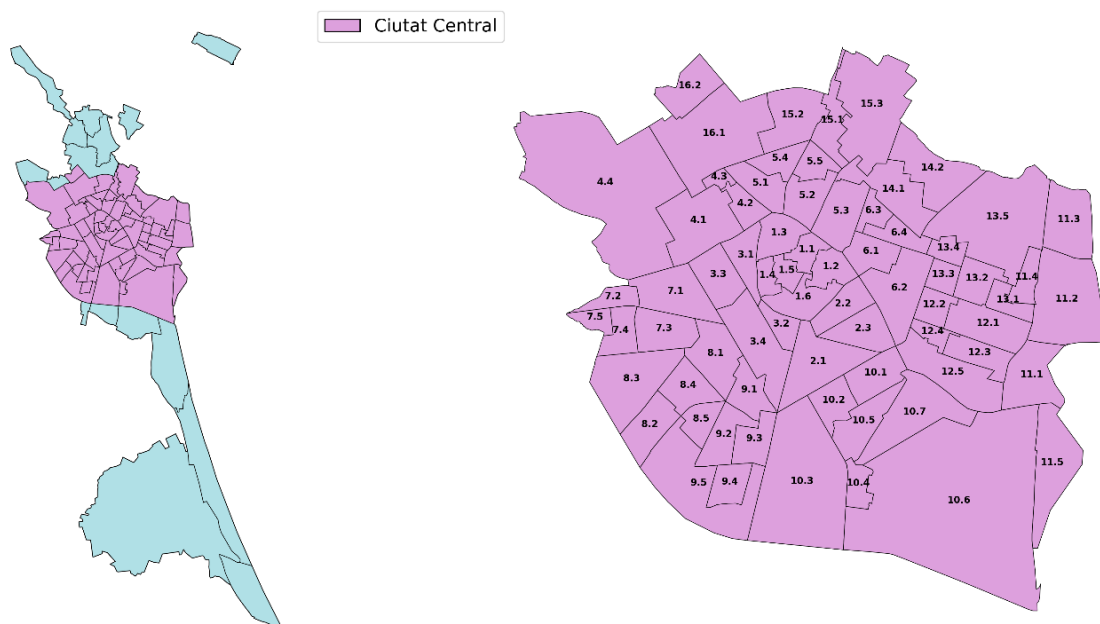
Debido a su extensión, al elevado número de habitantes y a la diversidad poblacional presente en la ciudad, las diferentes unidades geográficas que conforman Valencia presentan realidades y necesidades muy heterogéneas. De acuerdo con el *Plan Especial*

de *Directrices de Calidad Urbana* para los barrios de la ciudad (Ajuntament de València, 2018), se pueden identificar tres áreas territoriales distintas donde se debe abordar la problemática de calidad urbana mediante procesos independientes.

- ***Els Pobles***: comprende las pedanías de *Els Pobles del Nord*, *El Pobles de l'Oeste*, y *Els Pobles del Sud*, asentamientos urbanos de origen agrícola dispersos en la huerta o en el Parque Natural de L'Albufera. Estas regiones cuentan con una amplia extensión superficial, pero están escasamente pobladas.
- ***El Perellonet***: lo conforman los desarrollos litorales de segunda residencia, pertenecientes a la pedanía de *Els Pobles del Sud*.
- ***Ciutat Central***: integrada por los centros históricos de la ciudad y los desarrollos urbanísticos posteriores del siglo XX. En conjunto, la integran 70 barrios ubicados en los 16 distritos principales.

Como se resaltó anteriormente, los ámbitos territoriales de *Els Pobles* y *El Perellonet* no se ajustan completamente a la definición propia de una ciudad, sino que podrían calificarse como zonas rurales. Por esta razón, los análisis presentados a lo largo de este proyecto se llevarán a cabo sobre los 70 barrios centrales que conforman la *Ciutat Central*. En la Figura 1.3 se ven representados los 87 barrios de Valencia, destacando los 70 que corresponden a la región objeto de estudio.

Figura 1.3.- División administrativa de la ciudad de Valencia



Fuente: Elaboración propia

1.3. Indicadores de huella de carbono

Las ciudades desempeñan un papel crucial en la lucha contra el cambio climático a nivel mundial, siendo tanto las principales emisoras de Gases de Efecto Invernadero (GEI) como fuentes de soluciones climáticas innovadoras. Se calcula que el 70 por ciento de las emisiones globales de GEI relacionadas con la energía provienen de las ciudades, valor que seguirá en aumento como consecuencia del éxodo rural que se está produciendo del campo a las áreas urbanas (Mukim & Roberts, 2023; Standen, 2014).

De manera simultánea, las ciudades están desarrollando e implementando estrategias innovadoras para combatir el cambio climático, promoviendo el desarrollo sostenible y aumentando su resiliencia climática mientras reducen sus emisiones. No obstante, para maximizar su impacto global, se requiere de un estándar universal que permita a las ciudades medir sus emisiones e identificar las formas más efectivas de mitigarlas.

El Protocolo Global para Inventarios de Emisión de Gases de Efecto Invernadero a Escala Comunitaria (*Global Protocol for Community-Scale Greenhouse Gas Emission Inventories*, GPC) proporciona a las ciudades y gobiernos locales un marco sólido, transparente y aceptado a nivel universal para reconocer, calcular e informar de manera sistemática los Gases de Efecto Invernadero generados en las áreas urbanas. Este protocolo abarca tanto las emisiones producidas dentro de los límites de la ciudad como las producidas fuera de estos límites debido a las actividades que se desarrollan en la ciudad (Greenhouse Gas Protocol, 2022).

Esta iniciativa, desarrollada inicialmente en 1998 por el Instituto de Recursos Mundiales o WRI, por sus siglas en inglés (World Resources Institute, 2024), y el Consejo Mundial Empresarial para el Desarrollo Sostenible (WBCSD, 2024), es una alianza multipartita de empresas, organizaciones no gubernamentales (ONGs), gobiernos mundiales y otras entidades. Es la herramienta internacional más empleada para el cálculo y comunicación del Inventario de Emisiones.

Para elaborar el Inventario de Emisiones, se debe establecer un límite que identifique los gases, las fuentes de emisión, el área geográfica y el período de tiempo cubiertos por un inventario de GEI (Greenhouse Gas Protocol, 2022). El límite de inventario está diseñado para proporcionar a la ciudad que lo elabora una comprensión integral de dónde provienen las emisiones y de las áreas donde pueden tomar medidas. Específicamente, se delimita:

- **Límite geográfico.** Las ciudades deben establecer y mantener en el tiempo un perímetro geográfico que defina la extensión espacial del inventario. Este límite puede corresponder a un distrito concreto de la ciudad, una combinación de divisiones administrativas o cualquier otra entidad geográfica discernible.
- **Periodo de tiempo.** El inventario debe comprender un periodo de 12 meses.
- **Gases de Efecto Invernadero.** Las ciudades han de recoger en el inventario las emisiones de los siete gases contemplados como GEI en el Protocolo de Kioto, cuyo segundo periodo de compromiso finalizó en el año 2020 (United Nations Climate Change, n.d.). Estos gases son: dióxido de carbono (CO₂), metano (CH₄), óxido de nitrógeno (N₂O), hidrofluorocarbonos (HFCs), perfluorocarbonos (PFCs), hexafluoruro de azufre (SF₆) y trifluoruro de nitrógeno (NF₃). Estudios recientes concluyen que el CO₂ constituye el 90% de los inventarios urbanos de GEI (Lombardi et al., 2018).
- **Fuentes de emisión.** Las emisiones de GEI provenientes de las actividades de la ciudad se organizan en seis sectores principales, los cuales, a su vez, se dividen en diferentes subsectores. La Tabla 1.2 recoge esta información.

Con esta definición de sectores y subsectores, los Gases de Efecto Invernadero emitidos en una ciudad también pueden clasificarse según su alcance, debido a que las actividades que ocurren dentro de la ciudad pueden generar emisiones tanto dentro como fuera de sus límites geográficos. Para discernir este tipo de emisiones, el GPC define tres categorías o alcances en función del lugar donde se originan (Greenhouse Gas Protocol, 2022).

1. **Alcance 1.** Recoge las emisiones de GEI generadas por fuentes situadas dentro de los límites de la ciudad.
2. **Alcance 2.** Registra las emisiones de GEI producidas por el uso de energía, calor y/o sistemas de refrigeración suministrados a través de redes dentro de los límites de la ciudad.
3. **Alcance 3.** Abarca las emisiones restantes de GEI que se generan fuera de los límites de la ciudad, como consecuencia de las actividades llevadas a cabo dentro de dichos límites urbanos.

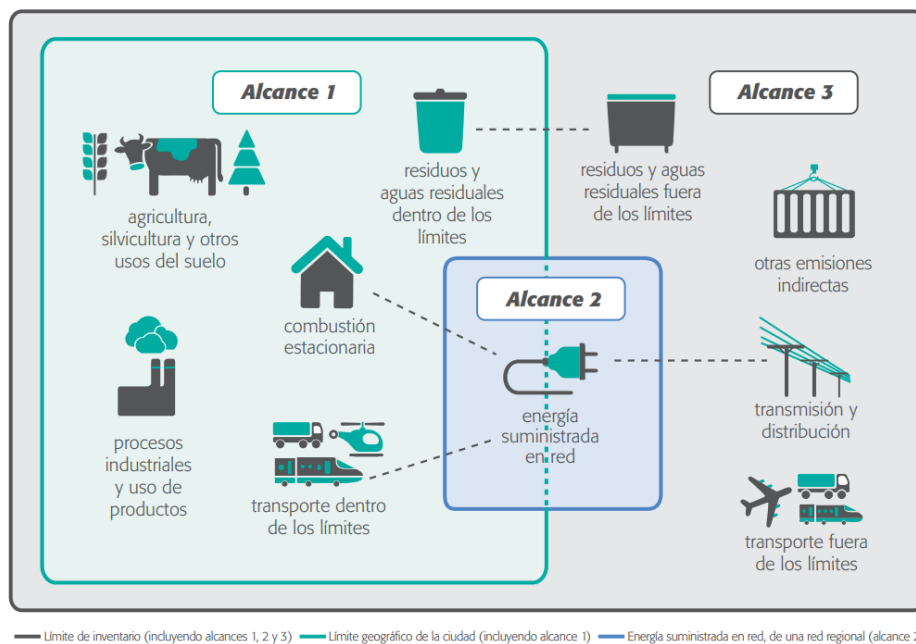
La Figura 1.4 presenta un esquema que resume el contenido de cada alcance.

Tabla 1.2.- Sectores y subsectores de las emisiones de GEI de una ciudad

Sector	Subsector
Energía estacionaria	Edificios residenciales
	Edificios e instalaciones comerciales e institucionales
	Industrias manufactureras y de la construcción
	Industrias de energía
	Actividades agrícolas, de silvicultura y de pesca
	Fuentes no especificadas
	Emisiones fugitivas provenientes de la minería, el procesamiento, el almacenamiento y el transporte de carbón
	Emisiones fugitivas provenientes de los sistemas de petróleo y gas natural
Transporte	Por carretera
	Ferroviario
	Navegación marítima, fluvial y lacustre
	Aviación
	Fuera de carretera
Residuos	Disposición de residuos sólidos
	Tratamiento biológico de residuos
	Incineración y quema a cielo abierto
	Tratamiento y vertido de aguas residuales
Procesos industriales y uso de productos	Procesos industriales
	Uso del producto
Agricultura, silvicultura y otros usos del suelo	Ganadería
	Suelo
	Fuentes agregadas y emisiones procedentes de fuentes del suelo distintas al CO ₂
Otras emisiones de alcance 3	

Fuente: *Greenhouse Gas Protocol* (2022)

La huella de carbono (HC) cuantifica las emisiones de Gases de Efecto Invernadero (GEI), expresadas como equivalente de dióxido de carbono o CO_{2e}, provenientes de las actividades humanas directas e indirectas en áreas urbanas (Lombardi et al., 2017). Los equivalentes de CO_{2e} son una medida de comparación del efecto de todos los GEI sobre el clima (Brander & Davis, 2012). Al convertir todas las emisiones a una unidad común, se pueden comparar sus impactos e incluir los efectos de todos los contaminantes en una única métrica. Los siete GEI previamente mencionados se convierten a CO_{2e} empleando el índice GWP (*Global Warming Potential* o *Potencial de Calentamiento Global*) (Miller et al., 2016).

Figura 1.4.- Fuentes y límites de las emisiones de GEI de la ciudad

Fuente: *Greenhouse Gas Protocol* (2022)

La información disponible, que servirá de base para los análisis llevados a cabo en este proyecto, ha sido recopilada considerando las categorías establecidas por el GPC en relación a la huella de carbono. Posteriormente, los datos serán organizados en base a los tres alcances que dictamina el GPC.

CAPÍTULO II: METODOLOGÍA

En este capítulo se ofrece una descripción detallada de los datos utilizados en la elaboración de este Trabajo Fin de Máster. Posteriormente, se expone de manera exhaustiva la metodología aplicada para la obtención de los resultados.

A partir del análisis de la autocorrelación espacial, se estudia si los datos presentan una relación en el espacio. Además, es posible comprender el comportamiento desde una perspectiva univariante de cada una de las variables incluidas en la base de datos. Por otro lado, con el fin de entender de una manera integral la información contenida en los datos, se propone un estudio multivariante mediante dos metodologías: el Análisis de Componentes Principales (PCA) como técnica de reducción de la dimensionalidad y varias técnicas de clustering, concretamente el algoritmo K-Means, el método de Ward y la regionalización. Finalmente, para relacionar los dos datasets disponibles, se presenta un estudio multivariante mediante la metodología Mínimos Cuadrados Parciales (PLS).

Todos los análisis aquí presentados han sido realizados mediante el lenguaje de programación Python (Van Rossum, 2024). Las librerías utilizadas han sido *Pandas* (The pandas development team, 2024), *GeoPandas* (Jordahl et al., 2020), *Shapely* (Gillies, 2024), *NumPy* (Harris et al., 2020), *Matplotlib* (Hunter, 2007), *Seaborn* (Waskom, 2021), *PySAL* (S. J. Rey & Anselin, 2007) y *Scikit-Learn* para los algoritmos de aprendizaje automático (Pedregosa et al., 2011).

2.1. Descripción de los datos

Para la elaboración del presente TFM, se han empleado dos bases de datos distintas sobre los que se realizarán los análisis pertinentes. Por un lado, para la realización del propio análisis estadístico espacial, se dispone de un dataset con variables que miden la huella de carbono para cada barrio. Por otro lado, se cuenta con un conjunto de datos que incluye numerosas variables socioeconómicas que definen las características de cada barrio.

2.1.1. Indicadores de huella de carbono

Según la clasificación establecida por el GPC, se han recogido una serie de indicadores que estiman la huella de carbono en cada barrio de la ciudad de Valencia. Roncero Tarazona et al. (2024) han calculado estos indicadores siguiendo la metodología descrita en su artículo "*Enabling Carbon Footprint accounting at neighbourhood level. Case of application to a Mediterranean compact city*". La recopilación de estos factores siguiendo la estructura administrativa actual de Valencia (dividida en 87 barrios, 70 de ellos centrales), facilitará la interpretación de los resultados obtenidos, proporcionando una perspectiva precisa y realista.

Específicamente, los análisis realizados en este proyecto se han llevado a cabo considerando dos conjuntos de indicadores relacionados:

- a. **Indicadores individuales.** Este grupo de variables se ha recopilado teniendo en cuenta las fuentes de emisión que dictamina el GPC, distinguiéndose un total de 24 indicadores. En la Tabla 2.1 se muestra una breve descripción de las emisiones que recoge cada uno de los indicadores.

Tabla 2.1.- Descripción de los indicadores individuales de huella de carbono

Indicador	Descripción
Food	Producción, transporte y consumo de alimentos
Manufact. products	Fabricación, transporte y consumo de productos elaborados
Trips	Viajes de larga distancia en tren, avión o coche privado
Waste	Tratamiento de residuos (vidrio, papel, plástico) extra municipales
Railway	Consumo eléctrico de metro y tranvía
Railway Losses	Pérdidas de energía por consumo eléctrico de metro y tranvía
Electricity Residential	Consumo de electricidad en los hogares (residencial)
Electricity Commercial	Consumo de electricidad en establecimientos comerciales
Electricity Industrial	Consumo de electricidad en la industria
NG Residential	Consumo de gas natural en los hogares (residencial)
NG Commercial	Consumo de gas natural en establecimientos comerciales
NG Industrial	Consumo de gas natural en la industria
Electricity Losses	Pérdidas por consumo de electricidad
NG Losses	Pérdidas por consumo de gas natural
Fossil Fuel Transport	Consumo de combustibles fósiles por los vehículos (transporte)
Electric Transport	Consumo de electricidad por los vehículos eléctricos (transporte)
FF Transport Losses	Pérdidas por consumo de combustibles fósiles por los vehículos
Building Material	Extracción, producción y transporte de materiales empleados en la construcción
On Site Construction	Proceso de construcción
Compost	Compostaje
Recycling	Reciclaje
Landfilling	Vertederos
Incineration	Incineración
Natural Sequestration	Absorción de carbono por las áreas verdes

b. Indicadores agrupados. Los seis indicadores agrupados que se han calculado y que recopilan la información de los indicadores individuales se basan en los tres alcances establecidos por el GPC. Se distinguen: *S1*, *S2*, *S3*, *S1S2* y *Total* (suma de *S1*, *S2* y *S3*).

Para el correcto desarrollo de este trabajo, se ha generado un archivo ESRI Shapefile (SHP) a partir del archivo Excel que contiene la información de estas variables. Este formato de archivo fue creado por la compañía desarrolladora de software de Sistemas de Información Geográfica ESRI (1998) y permite el almacenamiento y posterior tratamiento de datos espaciales.

Sistemas de Información Geográfica

Un Sistema de Información Geográfica (SIG o GIS, por sus siglas en inglés *Geographical Information System*) es una herramienta que permite el manejo de datos georreferenciados. Estos sistemas almacenan, analizan, gestionan y presentan datos que

vienen acompañados de una posición geográfica en la superficie terrestre. Para la utilización de estos sistemas, se debe disponer de un formato de archivo que contenga información acerca de la ubicación de los datos, como los archivos SHP o shapefile, uno de los formatos de datos más empelados en estos sistemas.

Un shapefile es un formato digital de almacenamiento de datos vectoriales que almacena la ubicación, la forma y los atributos asociados a las entidades geográficas. Un archivo SHP está compuesto por varios archivos que funcionan juntos, con las siguientes extensiones (ArcGis Desktop, n.d.):

- **Shape (extensión .shp).** Es el archivo principal que almacena la información geométrica de los elementos. Estas geometrías pueden ser puntos, líneas o polígonos.
- **Shape Index (extensión .shx).** Proporciona un índice de las posiciones de cada geometría que permite organizar y refinar las búsquedas dentro del archivo .shp.
- **dBase (extensión .dbf).** Recopila la base de datos en la que se registran los atributos asociados a cada elemento o geometría.
- **Project (extensión .prj).** Formato habitual, aunque no indispensable, que almacena el sistema de coordenadas utilizado para las geometrías.

Todos los análisis estadísticos espaciales aquí presentados han sido realizados empleando la información contenida en este conjunto archivos.

2.1.2. Variables socioeconómicas

Con el fin de realizar un análisis exploratorio de las características que definen los barrios de la ciudad de Valencia, así como de estudiar si existen relaciones entre estos indicadores socioeconómicos y los indicadores que miden la huella de carbono para cada barrio, se dispone de una segunda base de datos con 284 variables socioeconómicas para el año 2020. Los diferentes conjuntos de datos que forman este dataset han sido recopilados de la Oficina de Estadística (Ayuntamiento de Valencia, n.d.).

El conjunto de las 284 variables que se emplean se puede ver reflejado en el ANEXO II. Aquellas variables relacionadas con un conteo de población, de viviendas o de otro tipo, es decir, “número de...”, se han transformado a porcentaje con el fin de realizar un análisis más preciso y correcto.

2.2. Análisis de Componentes Principales

El *Análisis de Componentes Principales* o *Principal Component Analysis* (PCA) es una técnica multivariante de reducción de la dimensionalidad que permite simplificar la complejidad de espacios muestrales con gran cantidad de variables y/u observaciones, manteniendo la información relevante (Pearson, 1901). El uso del PCA desde una perspectiva de análisis exploratorio de datos facilita la comprensión del comportamiento de las variables empleadas y de sus interrelaciones, así como la identificación de individuos con características similares (Mardia et al., 2024).

Con una muestra de n individuos y p variables cuantitativas (X_1, X_2, \dots, X_p), se dispone de una matriz X de dimensiones $n \times p$, donde cada X_{ij} representa el valor del individuo i en la variable j . El objetivo de esta metodología es encontrar un número r de variables latentes o componentes principales (z_1, \dots, z_r) que recojan el comportamiento de las p variables originales para los n individuos de la muestra, teniendo en cuenta que $r < p$. Antes de proceder a la implementación del modelo PCA, es necesario estandarizar los datos asegurándose de que tengan media 0 y varianza 1 (Afifi et al., 2019).

Cada una de las z_i componentes principales se calcula como combinación lineal de todas las variables del conjunto p . De este modo, se calcula el *score* de cada componente para cada observación. Una de las características principales de las z_i es que la primera componente principal (z_1) siempre explica el mayor porcentaje de variabilidad presente en los datos y, progresivamente, las i componentes sucesivas explican un menor porcentaje de varianza. Además, todas las componentes obtenidas son ortogonales entre sí, es decir, son independientes y no están correlacionadas. La ecuación matemática que subyace en el cálculo de las z_i componentes principales se expone a continuación:

$$z_i = \phi_{1,i}X_1 + \phi_{2,i}X_2 + \dots + \phi_{p,i}X_p$$

donde $(\phi_{1,i}, \phi_{2,i}, \dots, \phi_{p,i})$ son los *loadings* del PCA y representan la importancia de cada variable X_j en cada componente i .

Para establecer el número óptimo de componentes principales a retener, se emplea una estrategia de visualización denominada Scree Plot, que muestra el porcentaje de varianza explicada para cada componente. Este gráfico ilustra cómo va disminuyendo dicho porcentaje a medida que se focaliza la atención en la componente siguiente.

2.3. Mínimos Cuadrados Parciales

La regresión por *Mínimos Cuadrados Parciales* o PLS, por sus siglas en inglés (*Partial Least Squares*), es una técnica de análisis multivariante que, al igual que el Análisis de Componentes Principales, se trata de un modelo basado en estructuras latentes. PLS combina características de la Regresión Lineal Múltiple (MLR) y del PCA, proporcionando una solución efectiva en situaciones donde se dispone de un gran número de variables X (predictoras) e Y (predichas) que pueden estar altamente correlacionadas.

Esta técnica se emplea para relacionar dos matrices de datos, X e Y, mediante un modelo lineal. A diferencia del modelo de regresión lineal tradicional, PLS permite modelar tanto la estructura de X como la de Y simultáneamente. Este modelo es especialmente útil cuando se dispone de grandes estructuras de datos complejos como los que se dispone en la actualidad, caracterizados por tener muchas variables correlacionadas y valores faltantes tanto en X como en Y (Vinzi et al., 2010).

La metodología fue desarrollada por Herman Wold (1979) como una alternativa a la Regresión Lineal Múltiple (MLR), donde se pudiesen analizar datos con numerosas variables explicativas X fuertemente correlacionadas, a la vez que modelar varias variables respuesta Y, también numerosas y correlacionadas. Este autor desarrolló en el año 1982 un algoritmo eficiente para la estimación de los parámetros, denominado NIPALS (*Non-linear Iterative Partial Least Squares*). Cabe destacar que PLS tiene la propiedad de que la precisión de los parámetros del modelo mejora conforme se incrementa el número de variables y observaciones en el modelo (S. Wold et al., 2001).

Las ecuaciones del modelo son las siguientes:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UC}' + \mathbf{F}$$

Donde \mathbf{X} es una matriz de dimensiones N observaciones y K variables explicativas e \mathbf{Y} una matriz de dimensiones N observaciones y M variables respuesta. Las matrices \mathbf{E} y \mathbf{F} recogen los errores del modelo. La matriz \mathbf{T} representa los scores en el espacio \mathbf{X} y se estiman como combinaciones lineales de las variables originales con los weightings w^* . Los weightings son los coeficientes de correlación entre las variables explicativas y los scores en el espacio \mathbf{Y} , representados en la matriz \mathbf{U} , de forma que:

$$T = XW^*$$

Finalmente, la matriz **C** recoge los loadings en el espacio **Y**, que representan las correlaciones entre estas variables y los scores de **X**; mientras que **P** es la matriz que almacena los loadings o direcciones de cada componente en el espacio **X**.

Para profundizar en esta técnica se pueden consultar los trabajos que conforman el origen del PLS (Geladi & Kowalski, 1986; H. Wold, 1979, 1982; S. Wold et al., 1984), así como artículos específicos que aplican esta técnica en el ámbito de la quimiometría (Kumar, 2021).

El modelo PLS presentado en este estudio se ha llevado a cabo empleando el software de análisis multivariante Aspen ProMV en su versión 10.0 (AspenTech, 2017). Este programa, desarrollado por la compañía estadounidense AspenTech, está diseñado para facilitar el manejo de grandes volúmenes de datos y analizar los factores que influyen en la variabilidad entre gran cantidad de variables correlacionadas en los procesos de fabricación de las empresas.

2.4. Análisis estadístico espacial

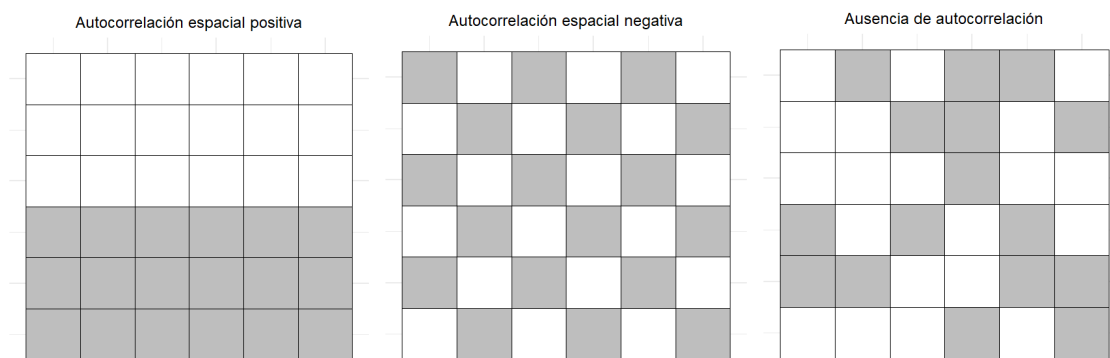
La utilización y el análisis de datos e información geográfica han experimentado un notable incremento en los últimos años, impulsados por la creciente disponibilidad de este tipo de datos espaciales. Esta información georreferenciada posee características distintivas que demandan el uso de herramientas estadísticas especializadas.

Específicamente, el Análisis Exploratorio de Datos Espaciales (ESDA) supera al Análisis Exploratorio de Datos (EDA) tradicional definido por Tukey debido a que, en el contexto de problemas geográficos, ayuda a determinar si los datos presentan un patrón espacial. El concepto que conecta formalmente la similitud geográfica y la estadística se denomina *autocorrelación espacial*, y surge como consecuencia de la existencia de una relación funcional entre lo que sucede en un punto específico del espacio y lo que sucede en otro lugar (Cliff & Ord, 1973; Paelinck, 1978). Como resultado, el valor de una variable en un área específico no se explica solo por factores internos, sino también por el valor de esa misma variable en regiones vecinas, lo que rompe el supuesto de independencia entre las observaciones (Moreno Serrano & Vayá-Valcarce, 2000).

La medición de cómo una misma variable se correlaciona en diferentes unidades espaciales contiguas puede resultar en una de estas tres situaciones, graficadas en la Figura 2.1 (Anselin, 1988):

- **Autocorrelación espacial positiva:** las unidades espaciales vecinas tienen valores similares, lo que refleja una tendencia al agrupamiento de las mismas.
- **Autocorrelación espacial negativa:** las unidades espaciales vecinas presentan valores muy diferentes. Esto indica una tendencia a la dispersión de las unidades espaciales.
- **Sin autocorrelación:** no se observa ninguna de las dos situaciones anteriores, lo que significa que los valores de las unidades espaciales vecinas son aleatorios.

Figura 2.1.- Patrones de autocorrelación espacial



Fuente: Elaboración propia

Arthur Getis (2007) destaca la importancia crucial de este tipo de análisis en el estudio de la econometría espacial. Este análisis permite identificar asociaciones entre unidades espaciales, formular hipótesis sobre estas relaciones e identificar outliers. Getis diferencia entre dos tipos de estadísticos, distinguiendo los mismos según si proporcionan respuestas globales o locales de los datos. En este trabajo se enfatiza en el estudio de los estadísticos de autocorrelación espacial global *I de Moran* (Moran, 1948) y *C de Geary* (Geary, 1954), mientras que para el estudio de la autocorrelación espacial local se focaliza la atención en el índice *I_i de Anselin* (Anselin, 1995).

Antes de proceder al estudio de los distintos estadísticos, es necesario que todas las representaciones de autocorrelación espacial incluyan una medida de la configuración topológica de las unidades espaciales, conocida como matriz de pesos espaciales.

2.4.1. Matriz de pesos espaciales

La matriz de pesos espaciales, comúnmente identificada por \mathbf{W} (“weight”), permite introducir la dependencia o vecindad entre las distintas regiones. Formalmente, esta matriz se representa como sigue:

$$\mathbf{W} = \begin{bmatrix} 0 & w_{1,2} & \cdots & w_{1,j} & \cdots & w_{1,N} \\ w_{2,1} & 0 & \cdots & w_{2,j} & \cdots & w_{2,N} \\ \vdots & \vdots & \ddots & \cdots & \cdots & \vdots \\ w_{i,1} & w_{i,2} & \vdots & 0 & w_{i,j+1} & w_{i,N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{N,1} & w_{N,2} & \vdots & w_{N,j} & \vdots & 0 \end{bmatrix}$$

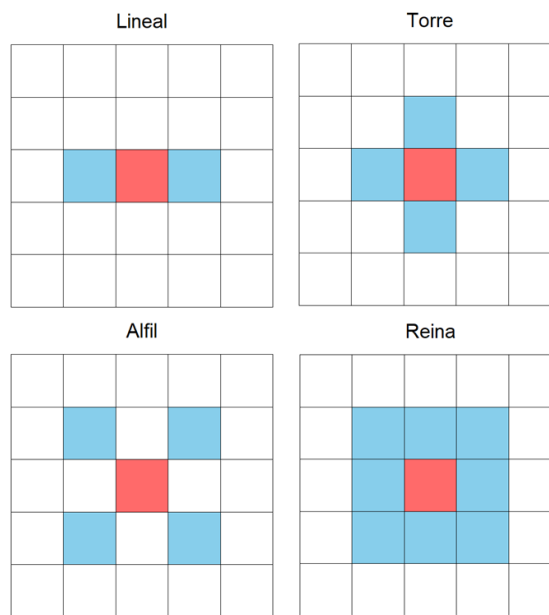
donde cada celda $w_{i,j}$ contiene un valor que representa el grado de interacción espacial entre cada par de observaciones o regiones i y j (Moreno Serrano & Vayá-Valcarce, 2000). Por tanto, será una matriz cuadrada de dimensiones $N \times N$ regiones.

Merece mencionar que no existe una definición unánimemente aceptada de la matriz \mathbf{W} , si bien se ha de cumplir en todo caso que los pesos sean no negativos y finitos (Anselin, 1980). No obstante, las matrices de pesos espaciales más empleadas son aquellas basadas en las distancias y las de contigüidad.

Matriz de pesos basados en contigüidad

El criterio más simple para el cálculo de la matriz de pesos espaciales es la contigüidad física de primer orden (Geary, 1954; Moran, 1948), que asigna los pesos en función de si los pares de ubicaciones comparten un límite o vértice común. En concreto, establece que $w_{i,j}$ es igual a 1 si las regiones i y j son físicamente adyacentes y 0 en caso contrario.

Los tipos de contigüidad más comúnmente empleados son la lineal, torre o *rook*, alfil o *bishop* y reina o *queen*, caracterizadas de este modo por su similitud con los movimientos en ajedrez, tal y como se visualiza en la Figura 2.2.

Figura 2.2.- Ejemplos de tipos de contigüidad física de primer orden

Fuente: Elaboración propia

Si bien las matrices de contigüidad se emplean en numerosas ocasiones por su simplicidad, una de sus principales limitaciones es que considera la adyacencia física como único determinante para establecer las interdependencias entre regiones (Moreno Serrano & Vayá-Valcarce, 2000).

Matriz de pesos basados en distancias

Una de las alternativas a las matrices de contigüidad son las matrices de pesos basados en distancias, propuestas inicialmente por Cliff & Ord (1973, 1981). Este tipo de matrices asignan mayores pesos a los pares de ubicaciones más cercanas y menores pesos a los pares más alejados. Dichas matrices son efectivas para capturar la autocorrelación espacial, dado que las ubicaciones próximas tienden a presentar valores similares (Anselin et al., 2013).

Un ejemplo muy común de este tipo de matrices son las que siguen el criterio K-vecinos más cercanos (*KNN*, por sus siglas en inglés), el cual asigna valores binarios a los pesos de las observaciones vecinas. Mediante este criterio, se debe establecer un valor predeterminado de k , que determina el número de observaciones más cercanas a una región. Si las unidades geográficas de las que se dispone son polígonos, para medir la distancia entre estos objetos se debe establecer un punto representativo para cada uno, denominado *centroide* o punto medio de la geometría (S. Rey et al., 2023).

Otro tipo de matrices de uso más avanzado, relacionadas directamente con las matrices basadas en distancias, son las matrices de pesos basados en kernel. Este tipo de matrices asignan los pesos empleando una función kernel que integra criterios de distancia y similitud, facilitando el manejo de datos heterogéneos, como puntos, líneas o polígonos. No obstante, requieren la especificación de más parámetros y pueden resultar costosas computacionalmente, pues combinan varios supuestos de otras matrices.

Los pesos de kernel reflejan la expectativa de que la proximidad espacial disminuya con la distancia. En esencia, los pesos asignados entre las observaciones i y j dependen de su distancia, modulada por una función kernel con características específicas. Todas estas funciones comparten la propiedad de disminuir con la distancia, incorporando así la Primera Ley de la Geografía de Tobler (“*Todas las cosas están relacionadas entre sí, pero las cosas más próximas en el espacio tienen una relación mayor que las distantes*”) (Tobler, 1970), aunque pueden hacerlo a diferentes ritmos según la distancia. Las funciones kernel más empleadas son la lineal, la polinomial, la sigmoide y la gaussiana.

Recientemente han ido surgiendo nuevos enfoques para mejorar la precisión en la determinación de las unidades espaciales vecinas en estudios específicos, que posteriormente se convierten en métodos de aplicación general, como la matriz de pesos en red (*Network Weight Matrix*) propuesta por Ermagun & Levinson (2018).

Finalmente, es importante mencionar que, comúnmente, se recurre a la estandarización de la matriz de pesos W dividiendo cada elemento $w_{i,j}$ por la suma total de la fila a la que pertenece $\left(\frac{w_{i,j}}{\sum w_{i,j}}\right)$, logrando que el peso total de todos los vecinos de una observación sea igual a uno. Esto se realiza con el fin de equilibrar la influencia total de cada región con sus vecinas, lo que provoca que la matriz deje de ser simétrica (Moreno Serrano & Vayá-Valcarce, 2000).

2.4.2. Autocorrelación espacial global

Los estadísticos de autocorrelación espacial global analizan el conjunto de observaciones espaciales o regiones de manera conjunta, considerando todas las unidades de análisis como un bloque a través de la media global del fenómeno estudiado (\bar{y}). Estos índices no capturan la heterogeneidad en el comportamiento del fenómeno estudiado, entendida esta como las variaciones locales del fenómeno que ocurren en cada región.

Por tanto, el propósito de estos indicadores es resumir la distribución espacial de los valores de cada unidad de análisis o región y obtener una cuantificación formal de cómo se desvían de un patrón aleatorio. Se trata de estadísticos para caracterizar un mapa en términos de su grado de agrupamiento y resumirlo, ya sea de forma visual o numérica.

Dos de los estadísticos de autocorrelación espacial global más extendidos en la comunidad académica son la *I de Moran* y la *C de Geary*. No obstante, antes de proceder a desarrollar los índices más empleados, se debe introducir un concepto fundamental: el *retraso o lag espacial*.

Retraso espacial

El concepto de *retraso o lag espacial* es una de las aplicaciones más comunes y directas de las matrices de pesos espaciales \mathbf{W} en el análisis espacial (S. Rey et al., 2023). Este operador captura el comportamiento de una determinada variable Y en el entorno inmediato de cada ubicación y suaviza los valores vecinos al multiplicarlos por la matriz \mathbf{W} , generando así una nueva variable Y_{sl} . De forma matricial, este concepto se puede expresar como:

$$Y_{sl} = \mathbf{W}Y$$

o en notación individual como:

$$y_{sl-i} = \sum_{j=1}^N w_{i,j} y_j$$

donde $w_{i,j}$ es la celda en \mathbf{W} sobre la i -ésima fila y j -ésima columna, capturando así la relación espacial entre las observaciones i y j . Si la matriz \mathbf{W} está estandarizada por filas, cada peso $w_{i,j}$ está acotado entre cero y uno, y el retraso espacial se convierte en un “promedio local” del valor de una variable Y en las ubicaciones vecinas de la región i .

El retraso espacial permite relacionar el comportamiento de Y en una región determinada con su patrón en el entorno inmediato. Posteriormente, los índices de autocorrelación espacial global emplearán cada región individual para construir medidas generales sobre la tendencia global del conjunto de datos.

I de Moran

El índice *I de Moran* es una medida estadística desarrollada por Alfred Pierce Moran (1948) que evalúa las variaciones de autocorrelación espacial entre valores vecinos más cercanos. Una característica del mismo es que solo tiene en cuenta los valores de las unidades de análisis o regiones determinadas a partir de la matriz de pesos \mathbf{W} , sin considerar el valor de la región central.

Este estadístico es el más empleado en los estudios medioambientales más recientes (Akhter & Afroz, 2024; Sun et al., 2024; Yang et al., 2024). La expresión para el cálculo de este estadístico se describe a continuación:

$$I = \frac{N}{\sum_{i=1}^N \sum_{j=1}^N w_{i,j}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} \cdot (y_i - \bar{y}) \cdot (y_j - \bar{y})}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

donde N es el número de observaciones, $(y_i - \bar{y})$ es el valor estandarizado de la variable de interés Y en la región i y $w_{i,j}$ es el valor correspondiente a la i -ésima fila y la j -ésima columna de la matriz de pesos espaciales \mathbf{W} .

Una interpretación gráfica de este estadístico se puede visualizar mediante el gráfico de Moran, a través del cual es posible explorar la naturaleza y la fuerza de la autocorrelación espacial, además del signo de la misma. En esencia, es un diagrama de dispersión tradicional en el que los valores originales de la variable de estudio Y se muestran frente a su retardo espacial Y_{s1} . Es relevante destacar que los valores de la variable Y suelen estandarizarse restándoles su media para facilitar la interpretación de los valores por encima o por debajo de la misma (Rey et al., 2023).

La cuantificación numérica del estadístico *I de Moran* permite plantear hipótesis de inferencia estadística sobre el grado de aleatoriedad espacial de los valores de la variable objetivo Y . En concreto, la hipótesis nula (H_0) asume que los valores están distribuidos de manera aleatoria en el espacio y su p-valor correspondiente representará la proporción de veces, bajo la hipótesis de aleatoriedad espacial, que los resultados son más extremos que el valor observado de la variable. Este p-valor proporciona una medida de la probabilidad de obtener el resultado observado bajo la hipótesis nula, lo que permite evaluar si los datos exhiben un patrón espacial significativo o si podrían haber ocurrido simplemente por azar.

$$\begin{cases} H_0: \text{presencia de aleatoriedad espacial} \\ H_1: \text{ausencia de aleatoriedad espacial} \end{cases}$$

El rango de valores sobre los que varía el *I de Moran* es $[-1, 1]$. Un valor de cero indica ausencia de autocorrelación espacial, un valor positivo (mayor que cero) sugiere autocorrelación espacial positiva, mientras que un valor negativo (menor que cero) refleja autocorrelación espacial negativa.

C de Geary

El estadístico *C*, propuesto por el estadístico Robert Charles Geary (1954), representa la relación de contigüidad entre las regiones. Específicamente, este estadístico se enfoca en cómo los valores de una variable en diferentes ubicaciones geográficas se relacionan entre sí, considerando especialmente las diferencias entre valores de ubicaciones adyacentes. Por tanto, a diferencia de otras medidas, es más sensible a las diferencias locales entre valores vecinos.

Este índice difiere del índice *I de Moran* en que el *C de Geary* sí incluye el valor de la unidad central en el análisis de autocorrelación. Además, no realiza ningún tipo de estandarización en los datos. La expresión matemática para su cálculo es la siguiente:

$$C = \frac{N}{2 \sum_{i=1}^N \sum_{j=1}^N w_{i,j}} \cdot \frac{\sum_{i=1}^N \sum_{j=1}^N w_{i,j} \cdot (y_i - y_j)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}$$

donde $(y_i - y_j)^2$ es la diferencia de cuadrados para los valores de la región *i* y la región *j*.

De manera similar al caso del índice *I de Moran*, se realiza inferencia estadística formulando la hipótesis nula (H_0) que asume la existencia de aleatoriedad espacial, calculándose el p-valor correspondiente.

De manera teórica, el estadístico *C de Geary* puede tomar valores dentro del intervalo $[0, 2]$ y funciona de manera inversa al estadístico *I de Moran*. Un valor igual a uno indica ausencia de autocorrelación espacial, un valor superior a uno refleja autocorrelación espacial negativa, mientras que un valor inferior a uno denota autocorrelación espacial positiva.

La Tabla 2.2; **Error! No se encuentra el origen de la referencia.** presenta un resumen de los valores teóricos de los índices de autocorrelación espacial global para

cada patrón espacial, donde se observa que los índices están inversamente relacionados. No obstante, resulta fundamental conocer que los estadísticos de autocorrelación espacial pertenecen al ámbito de la estadística deductiva. Esto implica que los resultados numéricos derivados de los análisis deben ser interpretados en el contexto de la hipótesis nula (H_0) y nunca se deben interpretar los valores del índice de manera aislada.

Tabla 2.2.- Patrones espaciales y valores teóricos de los estadísticos de autocorrelación espacial global

Estadístico	Rango	Autocorrelación espacial negativa	Ausencia de autocorrelación espacial	Autocorrelación espacial positiva
I de Moran	[-1, 1]	$I < 0$	$I = 0$	$I > 0$
C de Geary	[0, 2]	$C > 1$	$C = 1$	$C < 1$

En resumen, las medidas de autocorrelación espacial global proporcionan un resumen estadístico de todo el conjunto de datos, considerando todas las unidades espaciales de análisis como un todo. Estas medidas son una buena herramienta para resumir un conjunto de datos en un valor único que capture el grado de agrupación geográfica, si es positivo, o dispersión, si es negativo. Sin embargo, no señala áreas concretas donde se agrupan tipos específicos de valores (altos o bajos), que podrían considerarse valores atípicos. Para resolver esta limitación, se plantean los estadísticos de autocorrelación espacial local.

2.4.3. Autocorrelación espacial local

En el contexto del análisis estadístico espacial, la capacidad de analizar no solo las variaciones globales, sino también las locales, ha sido uno de los avances más significativos. Este enfoque representa un hito crucial en el desarrollo de este tipo de análisis, ya que los patrones globales a menudo ocultan las particularidades locales, que son esenciales para la comprensión y resolución de los problemas planteados.

En 1995, Luc Anselin (1995) introdujo la idea de los *Indicadores Locales de Asociación Espacial* o *LISA (Local Indicators of Spatial Association)*, por sus siglas en inglés. Estos indicadores permiten identificar subzonas donde se observa un agrupamiento o dispersión del fenómeno estudiado, asignando medidas de autocorrelación a cada unidad de análisis. De este modo, se puede identificar individualmente el nivel de agrupamiento o dispersión de cada región en relación con

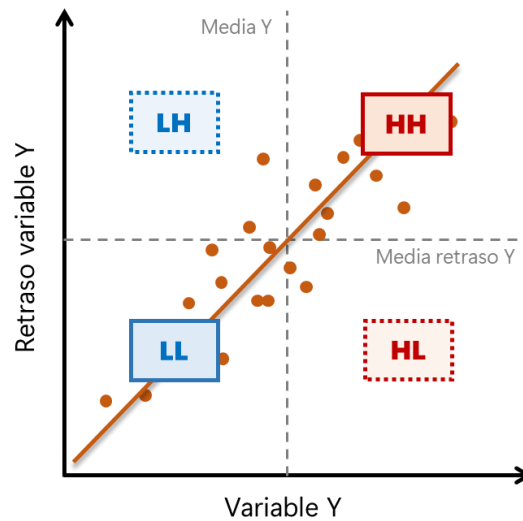
aquellas regiones que se hayan considerado próximas según la matriz de pesos W . Además, es posible localizar áreas con concentraciones inusuales de valores.

Los índices de autocorrelación espacial local no son estadísticas resumidas, sino puntuaciones que permiten comprender la estructura espacial de cada región. Algunos de ellos se relacionan directamente con los estadísticos globales, pues no son más que una descomposición en valores locales.

Para comprender el funcionamiento de estos estadísticos, se debe realizar una interpretación del gráfico de Moran (véase Figura 2.3). Previamente, los valores de la variable de estudio Y y del retraso espacial deben estandarizarse, en el primer caso restando la media (\bar{y}) y en el segundo caso normalizando previamente la matriz de pesos W . Tras este procedimiento, el gráfico de Moran puede dividirse en cuatro cuadrantes, donde cada región es clasificada en función de su propio valor y el de sus vecinos cercanos.

Los cuadrantes y sus respectivas comparaciones son la base del análisis, pues permiten identificar la presencia de agrupaciones (*clustering*) o valores atípicos en el fenómeno geográfico estudiado (Siabato & Guzmán-Manrique, 2019). Las agrupaciones o clústeres pueden presentar dos comportamientos opuestos, definidos en los cuadrantes I y III. En el cuadrante I (HH) se sitúan aquellas regiones que tanto ellas mismas como sus vecinos presentan valores altos por encima de la media (High-High). El cuadrante II (LL) agrupa las observaciones con valores bajos, tanto de la unidad central del análisis como de sus vecinos (Low-Low). Los cuadrantes II (LH) y IV (HL) determinan la presencia de valores atípicos, es decir, regiones que están rodeadas de valores vecinos opuestos a su valor, ya sea por encima (Low-High) o por debajo (High-Low).

Existen diversos índices capaces de medir la autocorrelación local, como los estadístico G_i y G_i^* de Getis y Ord (Getis & Ord, 1992; Ord & Getis, 1995), la C_i local de Geary o la I_i de Anselin (Anselin, 1995). Este último estadístico es el más extendido en la literatura por su facilidad de interpretación, pues muestra los resultados del índice para cada región i en un mapa.

Figura 2.3.- Gráfico de Moran teórico con autocorrelación positiva

Fuente: Elaboración propia a partir de S. Rey et al. (2023)

I_i de Anselin

El índice I_i de Anselin o I_i local de Moran (por ser la extensión del índice global de Moran) es un estadístico LISA que ayuda a comprender cómo cada unidad de análisis contribuye a la variabilidad total. En otras palabras, permite cuantificar la influencia de cada unidad en la autocorrelación global del área geográfica estudiada. Por lo tanto, mide la fuerza con la que los valores de una variable se agrupan en el espacio y detecta aquellos casos en los que una región y sus vecinos tienen valores similares (HH y LL) u opuestos (LH, HL), siendo estos últimos patrones lo esperado por azar. Para su cálculo se emplea la siguiente fórmula:

$$I_i = \frac{(y_i - \bar{y})}{m_2} \cdot \sum_{j=1}^N w_{i,j} \cdot (y_j - \bar{y}); \quad m_2 = \frac{\sum_{i=1}^N (y_i - \bar{y})^2}{N}$$

donde m_2 representa la varianza de la distribución de los datos.

La suma de las medidas I_i de Anselin para todas las unidades de análisis es proporcional al indicador global de asociación espacial I de Moran, considerándose este último como el promedio de las medidas locales (Siabato & Guzmán-Manrique, 2019).

Del mismo modo que se realizó para los indicadores globales, se plantean hipótesis de inferencia estadística que permitan identificar qué valores del estadístico son estadísticamente significativos para, posteriormente, clasificar los clústers en su etiqueta correspondiente (HH, LL, LH o HL).

2.5. Métodos de agrupación

Los métodos de agrupación o clustering son técnicas de aprendizaje no supervisado que se realizan con el objetivo de encontrar patrones o estructuras en los datos sin tener etiquetas o grupos predefinidos (Friedman, 2001). Estas técnicas de análisis multivariante tienen en cuenta las posibles relaciones entre variables, por lo que resultan de gran utilidad para realizar agrupaciones entre distintas observaciones con características similares.

El proceso de agrupación es un método fundamental de análisis geográfico, pues permite simplificar una amplia área geográfica identificando similitudes entre diversas dimensiones. Los grupos formados a partir de estas características similares se conocen como *clústeres*.

Específicamente, en la ciencia de datos geográficos, la agrupación se utiliza ampliamente para proporcionar información sobre la estructura geográfica de datos multivariantes espaciales complejos (S. Rey et al., 2023). En el contexto de cuestiones explícitamente espaciales, surge un concepto relacionado, la *región*. Una *región* se asemeja a un *clúster* en que todos sus miembros están agrupados y que debe proporcionar una versión simplificada de los datos originales contenidos en ella. Para que una región sea útil analíticamente, los miembros de la misma deben ser más similares entre sí que con los miembros de otras regiones. No obstante, las regiones son más complejas que los clústeres, ya que incorporan información adicional sobre la ubicación de sus miembros y describen un área geográfica definida. El proceso de creación de regiones se denomina regionalización (Duque et al., 2007).

2.5.1. Clustering de unidades geográficas

El análisis clúster geodemográfico es una técnica de agrupamiento en la que las observaciones se corresponden con áreas geográficas (Webber & Burrows, 2018), lo que permite que los resultados puedan visualizarse en un mapa. Esta técnica utiliza algoritmos de agrupamiento multivariante de la misma manera que en el caso general de clustering. La diferencia radica en que la etiqueta asignada a cada clúster se representa en un mapa, facilitando así la interpretación de los resultados.

Para el desarrollo de los análisis correspondientes, en el presente estudio se emplean dos algoritmos de agrupamiento: K-Means, perteneciente a la técnica de clustering no jerárquico, y el método de Ward, incluido en el clustering jerárquico aglomerativo. Estos

algoritmos han sido seleccionados por su amplio uso y popularidad en la literatura, así como por su sencilla aplicación en este contexto geográfico (Lamb et al., 2020; Varghese et al., 2013). Además, se busca realizar una comparativa entre ellos para observar sus diferencias.

K-means

El algoritmo K-means (Lloyd, 1982) es uno de los enfoques más empleados para agrupar conjuntos de datos. Este algoritmo agrupa las N observaciones en un número K predefinido de clústeres, de manera que cada observación se encuentre más próxima a la media de su propio clúster que a la de cualquier otro.

Este algoritmo, inicialmente propuesto por Lloyd (1982), sigue los siguientes pasos (Sammut & Webb, 2010):

1. Seleccionar aleatoriamente los K centroides iniciales
2. Asignar cada observación i al centroide más cercano, creando K clústeres iniciales.
3. Calcular los nuevos centroides a partir de las medias de las distancias euclídeas de las observaciones incluidas y actualizar los clústeres.
4. Repetir los pasos 2 y 3 hasta que los centroides ya no cambien significativamente entre iteraciones o hasta que se alcance el criterio de parada.
5. Devolver las observaciones con los clústeres asignados y los centroides finales, que representan los centros de los K clústeres.

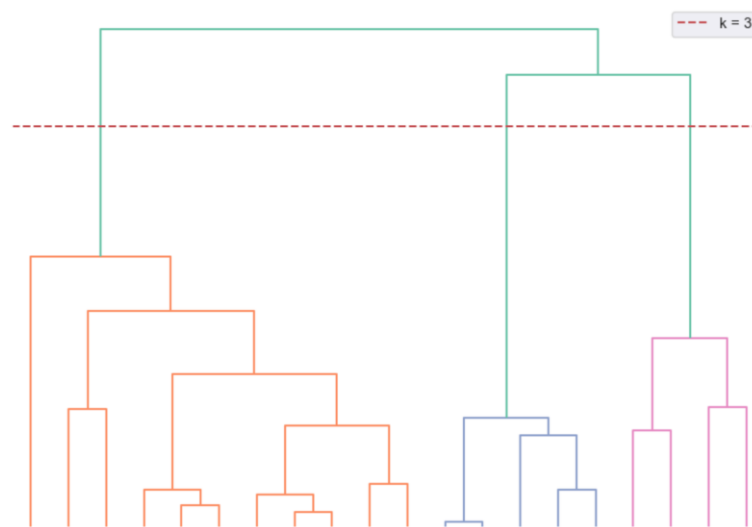
Método de Ward

Otro de los métodos de agrupación más empleados es el *Agglomerative Hierarchical Clustering* o AHC. El algoritmo AHC funciona construyendo una jerarquía de soluciones de agrupación ascendente, comenzando definiendo cada observación como un único clúster ($K = N$). En cada paso, el algoritmo encuentra los dos grupos más cercanos según una distancia predefinida entre clústeres. Estos dos clústeres se fusionan en un nuevo clúster, ascendiendo gradualmente en la jerarquía. El proceso finaliza cuando todas las observaciones se encuentran agrupadas en un solo clúster ($K = 1$). El término “aglomerativo” deriva del enfoque ascendente del algoritmo.

Para el presente estudio, se sigue el criterio de Ward con la distancia euclídea para escoger los pares de conglomerados a fusionar, pues es uno de los criterios que mejores resultados proporciona.

La ventaja de los métodos de agrupamiento jerárquicos es que permiten la creación de un dendograma (véase Figura 2.4), que no es más que una representación en un árbol de clasificación de las agrupaciones entre observaciones.

Figura 2.4.- Dendograma teórico seleccionando $k = 3$ clústeres



Fuente: Elaboración propia¹

Selección del número óptimo de clústeres

Como se ha comentado, la naturaleza de estos dos algoritmos requiere seleccionar previamente el número K de clústeres que se desean formar. Una de las metodologías más empleadas para establecer el número óptimo de clústeres es el coeficiente de Silhouette (Kaufman & Rousseeuw, 1990).

El coeficiente de Silhouette es una métrica que se emplea para evaluar la calidad del agrupamiento obtenido mediante un algoritmo de clustering. Se basa en la idea de que los puntos dentro de un mismo clúster deberían estar más cerca entre sí que a los puntos de otros clústeres. El principal objetivo es identificar el número óptimo de agrupamientos. Un valor más elevado para este indicador refleja un número de clústeres más deseable.

¹ En lo sucesivo, y con el fin de no resultar reiterativo, siempre que se omita la fuente se considera de elaboración propia.

El valor del coeficiente de Silhouette para una observación i , denotado por $s(i)$, se calcula como (Kaufman & Rousseeuw, 1990):

$$s(i) = \frac{b(i) - a(i)}{\max \{a(i), b(i)\}}$$

donde $a(i)$ representa el promedio de las distancias de la observación i con las demás observaciones del clúster al que pertenece y $b(i)$ es la distancia mínima a otro clúster que no es el mismo en el que está la observación i . Ese clúster es la segunda mejor opción para i .

El valor de $s(i)$ se encuentra comprendido en el rango $[-1,1]$ y, combinando los valores de $a(i)$ y $b(i)$, se obtiene:

$$s(i) = \begin{cases} 1 - \frac{a(i)}{b(i)}, & \text{si } a(i) < b(i) \\ 0, & \text{si } a(i) = b(i) \\ \frac{b(i)}{a(i)} - 1, & \text{si } a(i) > b(i) \end{cases}$$

- Si $s(i) \approx 1$, la observación i tiene gran similitud con su propio clúster, por lo que estará correctamente asignada.
- Si $s(i) \approx 0$, la observación i está en la frontera entre dos clústeres, es decir, existen dos clústeres superpuestos.
- Si $s(i) \approx -1$, la observación i se asemeja más a su clúster vecino, por lo que estará mal asignada y debería ser asignada a su clúster vecino.

Finalmente, se puede calcular el coeficiente de Silhouette como el promedio de todos los $s(i)$ para todas las observaciones del conjunto de datos.

2.5.2. Regionalización

La regionalización es un tipo espacial de agrupación o clustering que impone en su metodología una restricción espacial a las observaciones. Al ser un algoritmo explícitamente espacial de agrupación, proporciona información muy valiosa acerca de la estructura espacial de las relaciones estadísticas multivariantes entre las variables (Duque et al., 2007). De este modo, las “regiones” que se generan se consideran clústeres internamente conectados, y cada uno de los miembros de una región (agrupados según su similitud estadística) deberá estar geográficamente ubicado dentro de los límites de esa

región. Su principal objetivo es agrupar observaciones que son similares en sus atributos estadísticos, pero también en su ubicación espacial, consiguiendo que los clústeres formados no se encuentren fragmentados.

Como ocurría en el caso no espacial, existe una gran variedad de algoritmos de regionalización. No obstante, todos los métodos emplean un conjunto de atributos de entrada, así como una matriz binaria de pesos espaciales que representa la conectividad espacial entre las observaciones. Esta matriz de pesos espaciales establece la restricción espacial en este tipo de métodos (S. Rey et al., 2023).

Tal y como se mencionó en la explicación del análisis estadístico espacial, existen numerosos tipos de matrices que pueden ser calculadas para establecer la restricción espacial, como las de contigüidad de tipo alfil o *bishop*, torre o *rook*, reina o *queen*, así como las basadas en distancias, como KNN.

Finalmente, cabe destacar que no siempre es necesario aplicar un enfoque de regionalización a los datos. No obstante, en algunos casos puede proporcionar información adicional sobre la estructura espacial de los datos que los métodos de agrupación tradicionales no logran capturar.

CAPÍTULO III: RESULTADOS

Una vez se ha presentado la metodología empleada para el desarrollo del presente Trabajo Fin de Máster y se han descrito las bases de datos, en el presente capítulo se detallan los resultados obtenidos divididos en tres secciones. La primera sección está centrada en analizar los indicadores individuales de huella de carbono, tanto desde una perspectiva espacial como desde una perspectiva multivariante mediante el Análisis de Componentes Principales (PCA). En la segunda sección se analiza la información contenida en los indicadores agrupados de huella de carbono, realizando de nuevo un análisis espacial y añadiendo técnicas de clustering y regionalización con el fin de contrastar los resultados. Finalmente, en la última sección se desarrolla la metodología de Mínimos Cuadrados Parciales (PLS) con el propósito de observar si existe una relación entre las variables socioeconómicas que definen cada barrio y los indicadores agrupados de huella de carbono.

3.1. Indicadores individuales de huella de carbono

El primer análisis de los indicadores de huella de carbono se ha realizado para los 24 indicadores individuales. Por un lado, se detalla el análisis estadístico espacial de cada indicador. A continuación, se procede a implementar el Análisis de Componentes Principales (PCA) como base del análisis multivariante.

Antes de proceder con los análisis correspondientes, se ha realizado un análisis descriptivo de las variables para identificar si alguna de ellas era constante o casi constante calculando su desviación típica. Esto resulta fundamental, ya que si la variación de alguna variable es meramente ruido, al escalar las variables dicho ruido tendrá una mayor influencia en el análisis (Wold et al., 1987). La Tabla 3.1 muestra un resumen estadístico de las variables.

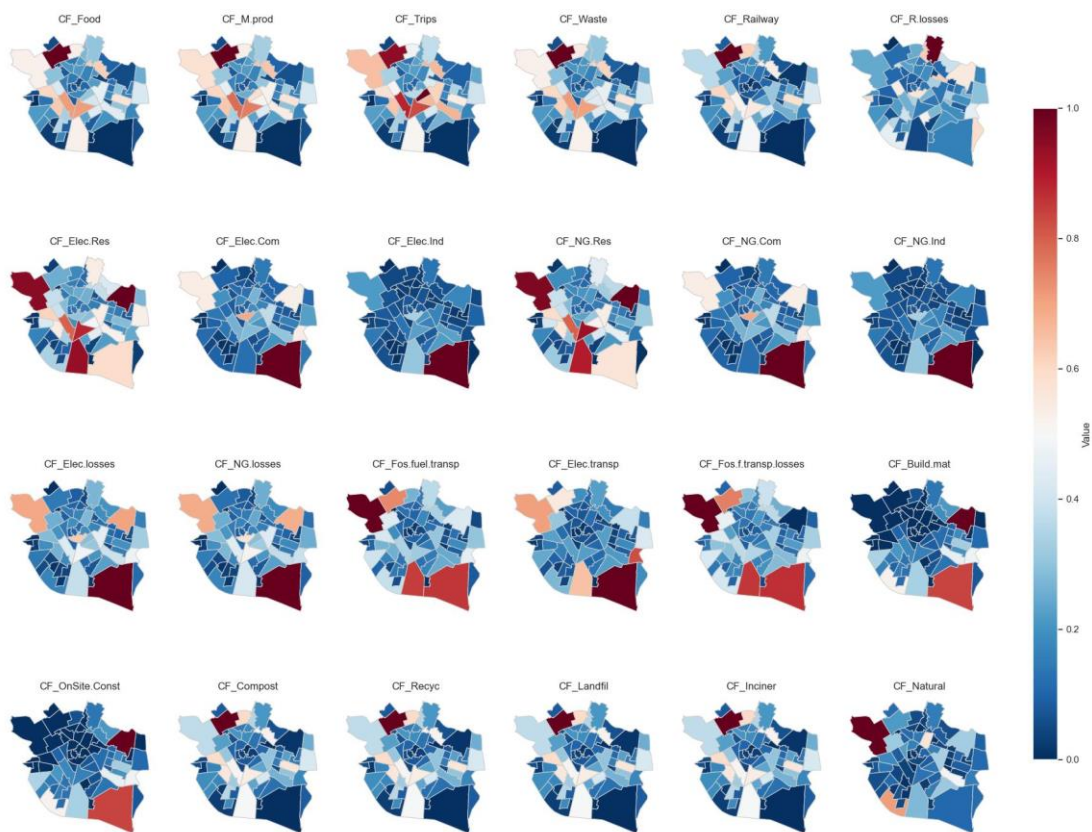
Tabla 3.1.- Resumen estadístico de los indicadores individuales

Indicador	Media	Desv. Típica	Mín.	Mediana	Máx.
Food	7952,35	5276,50	1717,03	6510,06	27273,48
M.prod	6196,83	4094,67	1268,86	5056,07	20034,96
Trips	423,90	298,27	73,01	328,47	1304,36
Waste	1056,98	693,77	233,08	888,37	3575,45
Railway	2,47	1,71	0,61	2,00	9,47
Railway.Losses	0,38	0,26	0,03	0,31	1,46
Elec.Resid	1412,01	1120,86	154,35	1101,62	4795,89
Elec.Comm	1996,54	2052,36	195,32	1416,09	12329,01
Elec.Indust	117,95	160,00	9,28	76,93	1226,30
NG.Resid	1156,68	990,54	131,55	828,01	4244,67
NG.Comm	1325,45	1362,51	129,66	940,10	8184,89
NG.Indust	243,81	330,72	19,18	159,01	2534,84
Elec.Losses	4548,83	3838,95	584,96	3597,56	21221,58
NG.Losses	157,60	135,46	20,58	119,00	760,88
Fos.Fuel.Transp	8570,88	7256,97	1646,44	6064,84	36993,33
Elec.Transp	43,19	42,78	7,43	27,61	232,07
Fos.F.Trans.Losses	1255,03	1093,62	0,00	902,72	5571,21
Build.Mat	462,66	719,40	0,00	289,70	3909,10
OnsSite.Const	63,16	98,21	0,00	39,55	533,65
Compost	140081,74	96734,63	35634,22	111830,94	543707,56
Recycling	65467,23	45208,95	16653,66	52264,21	254101,83
Landfilling	1244780,78	859593,83	316649,35	993741,21	4831441,34
Incineration	231265,63	159702,42	58829,72	184625,43	897624,97
Natural.Seq	43,31	56,42	0,44	28,22	336,22

Tras comprobar que ninguna variable toma valores casi constantes, se procede a detallar los análisis correspondientes.

Análogamente, se han realizado 24 mapas coropléticos (véase Figura 3.1), uno para cada indicador, con el fin de visualizar los valores obtenidos en cada barrio. Normalizando los datos en un rango entre 0 y 1, en color granate se observan los barrios con valores más elevados para las variables, mientras que en azul oscuro se muestran los valores más bajos, próximos a cero e incluso tomando el valor de cero.

Figura 3.1.- Mapas coropléticos de cada indicador individual coloreados por los valores obtenidos en cada barrio



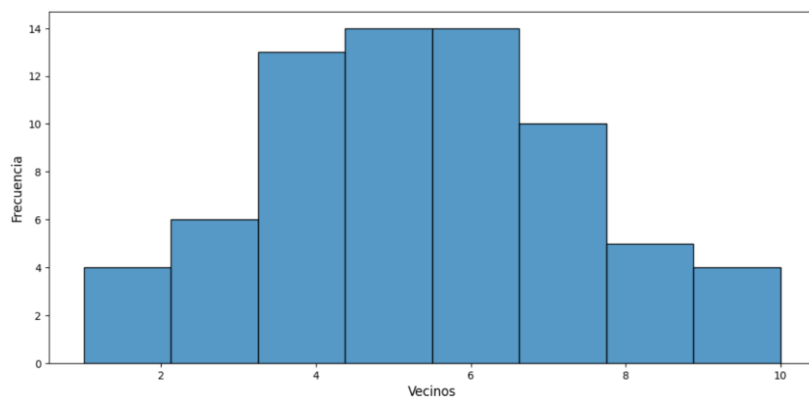
3.1.1. Análisis estadístico espacial

Para proceder con la implementación del análisis estadístico espacial, previamente se debe hallar la matriz de pesos espaciales \mathbf{W} . En este caso, en primer lugar se han estimado los centroides de cada barrio, pues son imprescindibles para poder calcular las distancias entre los polígonos irregulares y calcular posteriormente la matriz \mathbf{W} .

A continuación, la matriz de pesos espaciales que se ha calculado para realizar el análisis estadístico espacial es la matriz de pesos basados en kernel gaussiano. Para su cálculo, se han tenido en cuenta los siguientes parámetros:

1. **Vecinos cercanos (k).** En este parámetro se fija el número de vecinos cercanos para cada barrio. Si un barrio i no pertenece al número de vecinos k , ese barrio tomará un valor de 0 en la matriz \mathbf{W} . Para establecer el parámetro k óptimo, se ha calculado la matriz de adyacencia A según la contigüidad de tipo Reina. Como se observa en la Figura 3.2, cuatro barrios tienen un número de vecinos cercanos, según la contigüidad de tipo Reina, igual a 10. En consecuencia, con el fin de no dejar fuera del análisis barrios vecinos con capacidad de influencia, se determina un valor fijo de $k = 10$.

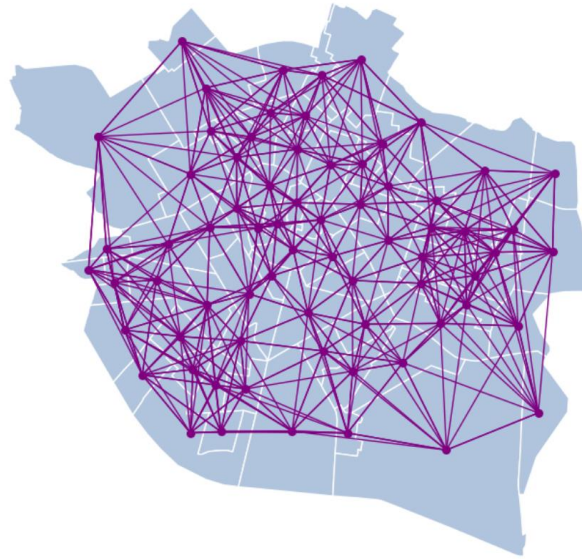
Figura 3.2.- Número de vecinos cercanos para cada barrio según la contigüidad de tipo Reina



2. **Función kernel ($K(z)$).** Tras haber analizado la distribución de las distancias entre los 70 barrios, mediante el test de normalidad Shapiro-Wilk se ha comprobado que, para un nivel de significación del 99%, los 70 barrios cumplen con la hipótesis nula (H_0) de normalidad. En consecuencia, la función kernel escogida para el cálculo de la matriz \mathbf{W} es la gaussiana.
3. **Ancho de banda (h).** Este parámetro especifica la distancia desde cada barrio sobre la que se aplica la función kernel. Puede ser fijo, si todas las observaciones emplean la misma distancia, o adaptativo. En este caso, se ha seleccionado un ancho de banda adaptativo, ya que la densidad de los barrios no es uniforme.

Con estos datos, se calcula finalmente la matriz de pesos espaciales basados en kernel gaussiano, de dimensiones 70×70 . En la Figura 3.3 se ven representados los 70 barrios y sus conexiones con los vecinos cercanos según la matriz de pesos espaciales estimada.

Figura 3.3.- Representación de las conexiones de los 70 barrios de Valencia según la matriz de pesos basada en kernel gaussiano con $k = 10$



La Figura 3.4 muestra los pesos asignados a cada barrio central (marcado con un punto negro). El color azul para los barrios refleja que tienen un peso igual a cero para esa región, es decir, no se consideran vecinos.

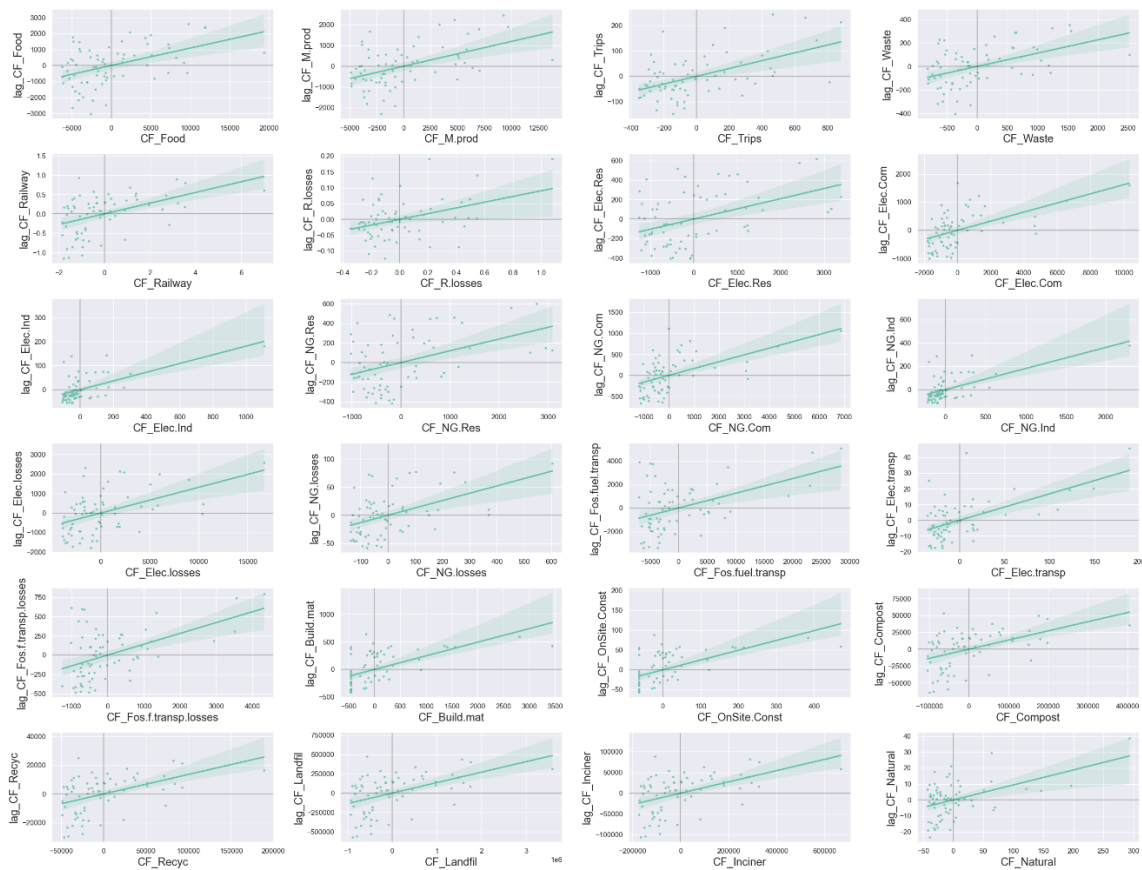
Figura 3.4.- Mapas de los vecinos cercanos asociados a cada barrio empleando la matriz de pesos basada en kernel gaussiano con $k = 10$



Autocorrelación espacial global

Una vez se ha calculado la matriz de pesos espaciales, a continuación se realiza el gráfico de Moran para cada variable con el objetivo de observar la autocorrelación espacial entre los valores originales de cada variable y su retardo espacial. Como se observa en la Figura 3.5, en todos los casos existe una autocorrelación espacial positiva entre cada variable y su retardo espacial.

Figura 3.5.- Gráficos de Moran para los 24 indicadores individuales



Para analizar cuantitativamente esta relación y verificar si esta autocorrelación espacial positiva es estadísticamente significativa, se han calculado los estadísticos de autocorrelación espacial global *I de Moran* y *C de Geary*. La Tabla 3.2 muestra, para cada variable, el valor de los índices y su p-valor asociado.

El cálculo de los p-valores asociados a los estadísticos que estima la librería *esda* de Python se realiza mediante simulación, con un número predeterminado de 999 permutaciones. Por tanto, con el fin de obtener siempre los mismos resultados, se ha fijado una semilla aleatoria.

Tabla 3.2.- Estadísticos I de Moran y C de Geary para los indicadores individuales

Indicador	I de Moran	p-valor (I)	C de Geary	p-valor (C)
Food	0,10931	0,364	0,850416	0,25
M.prod	0,11906	0,316	0,842273	0,186
Trips	0,15353	0,111	0,81359	0,082
Waste	0,11294	0,355	0,84673	0,224
Railway	0,13799	0,167	0,821919	0,12
Railway.Losses	0,09064	0,427	0,80808	0,09
Elec.Resid	0,10454	0,458	0,812597	0,086
Elec.Comm	0,16188	0,086	0,681215	0,005
Elec.Indust	0,18071	0,017	0,593154	0,002
NG.Resid	0,12004	0,314	0,801359	0,092
NG.Comm	0,16188	0,073	0,681215	0,005
NG.Indust	0,18071	0,01	0,593154	0,003
Elec.Losses	0,13140	0,217	0,726594	0,006
NG.Losses	0,13021	0,215	0,724896	0,007
Fos.Fuel.Transp	0,12559	0,256	0,7203	0,009
Elec.Transp	0,16812	0,077	0,661098	0,002
Fos.F.Trans.Losses	0,14080	0,162	0,701657	0,003
Build.Mat	0,24677	0,003	0,612714	0,001
OnSite.Const	0,24677	0,007	0,612714	0,001
Compost	0,13601	0,196	0,821289	0,137
Recycling	0,13601	0,174	0,821289	0,131
Landfilling	0,13601	0,197	0,821289	0,132
Incineration	0,13601	0,198	0,821289	0,136
Natural.Seq	0,09387	0,435	0,766429	0,073

Como se observa en la Tabla 3.2, el índice *I de Moran* solo ha resultado estadísticamente significativo, considerando un nivel de significación del 90%, para las variables “*Electricity Commercial*”, “*Electricity Industrial*”, “*NG Commercial*”, “*NG Industrial*”, “*Electric Transport*”, “*Building Material*” y “*On Site Construction*”. Dado que la mayoría de las variables no muestran significancia estadística, esto sugiere que no existe una tendencia general de agrupamiento en toda el área conjunta de los 70 barrios de Valencia. En otras palabras, la distribución espacial global de estas variables es aleatoria.

No obstante, si se analiza el índice *C de Geary*, el número de variables estadísticamente significativas se incrementa, incluyéndose “*Trips*”, “*Railway Losses*”, “*Electricity Residential*”, “*NG Residential*”, “*Electricity Losses*”, “*NG Losses*”, “*Fossil Fuel Transport*”, “*Fossil Fuel Transport Losses*” y “*Natural Sequestration*”.

Esta discrepancia entre ambos estadísticos se debe a que existen agrupamientos locales en regiones específicas, pero no existe una tendencia general en toda la Ciutat Central si se consideran las 24 variables. Es decir, se puede decir que existe una variación significativa en los valores de las observaciones adyacentes, aunque no hay un patrón global claro de clustering.

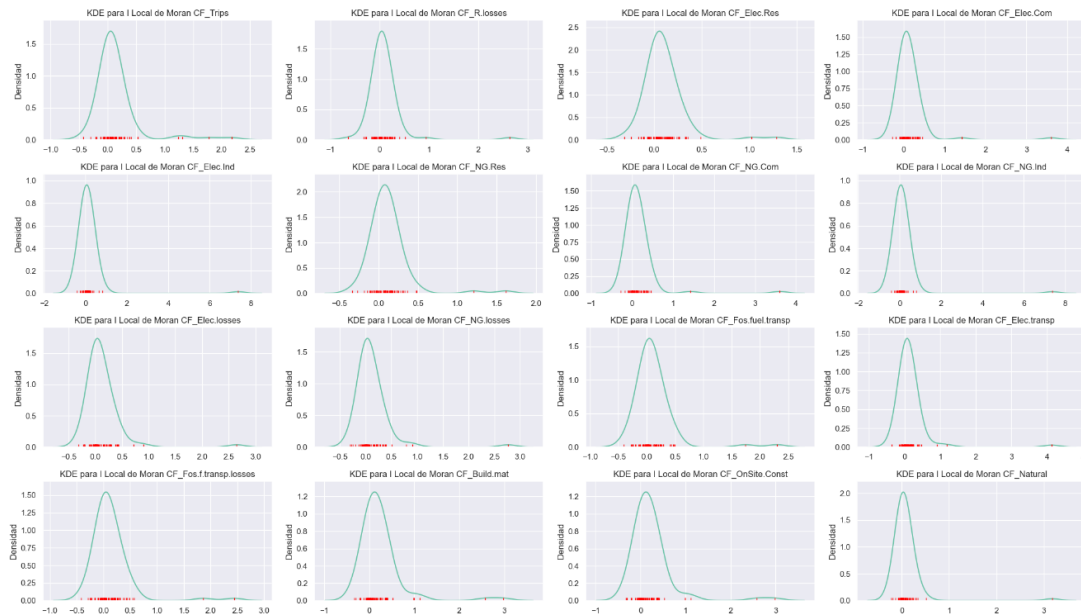
Al haberse encontrado resultados contradictorios entre la *I de Moran* y la *C de Geary*, esto permite guiar el análisis hacia una comprensión más detallada de las diferencias locales entre barrios. Por tanto, con el fin de comprender mejor la distribución de las variables para las que el *C de Geary* ha resultado estadísticamente significativo, se deben estudiar los agrupamientos concretos mediante los estadísticos de autocorrelación espacial local.

Autocorrelación espacial local

Tras haber analizado la autocorrelación espacial global para cada uno de los indicadores, seguidamente se estudiará en qué regiones concretas se agrupan tipos específicos de valores con ayuda del estadístico local *I_i de Anselin*.

En primer lugar, los gráficos Kernel Density Estimation o KDE mostrados en la Figura 3.6 permiten visualizar la distribución de los valores del índice *I_i de Anselin* para las variables que resultaron estadísticamente significativas para un nivel de significación del 90% en la *C de Geary*. De manera generalizada, en las gráficas se observa un ligero sesgo a la izquierda de la distribución del estadístico, lo que refleja de nuevo que existe autocorrelación espacial positiva. Relacionando estos resultados con los gráficos de Moran mostrados en la Figura 3.5, valores altos (bajos) de las variables estarán próximos a valores altos (bajos) en el conjunto de la ciudad de Valencia, por lo que podrán observarse regiones con clasificación HH (LL).

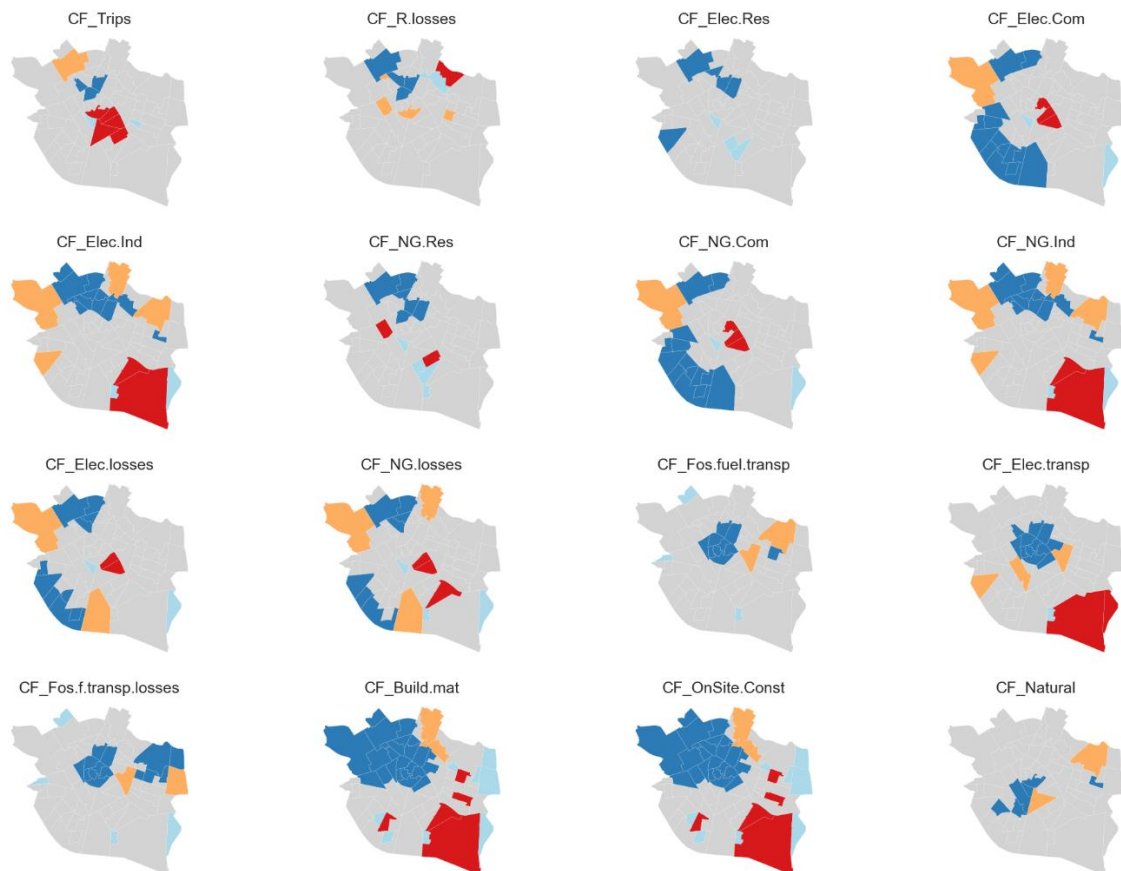
Figura 3.6.- Gráfico KDE del estadístico I_i de Anselin para cada uno de los indicadores individuales



No obstante, del mismo modo que antes, se debe cuantificar esta relación con el fin de analizar en qué regiones específicas esta autocorrelación es significativa y en cuáles existen valores atípicos. Para ello se recurre a la clasificación del gráfico de Moran explicada en el capítulo anterior, donde se dividió este gráfico en cuatro cuadrantes (HH, LL, LH y HL). Para observar los clústeres estadísticamente significativos en la I_i de Anselin para un nivel de confianza del 90%, se ha graficado un mapa de la ciudad de Valencia para cada variable, clasificando los mismos según el comportamiento presentado (véase Figura 3.7). De la información mostrada se deben destacar las regiones coloreadas en color rojo (HH) y azul (LL), las cuales se asocian con valores altos y bajos para cada uno de los indicadores y sus vecinos, respectivamente.

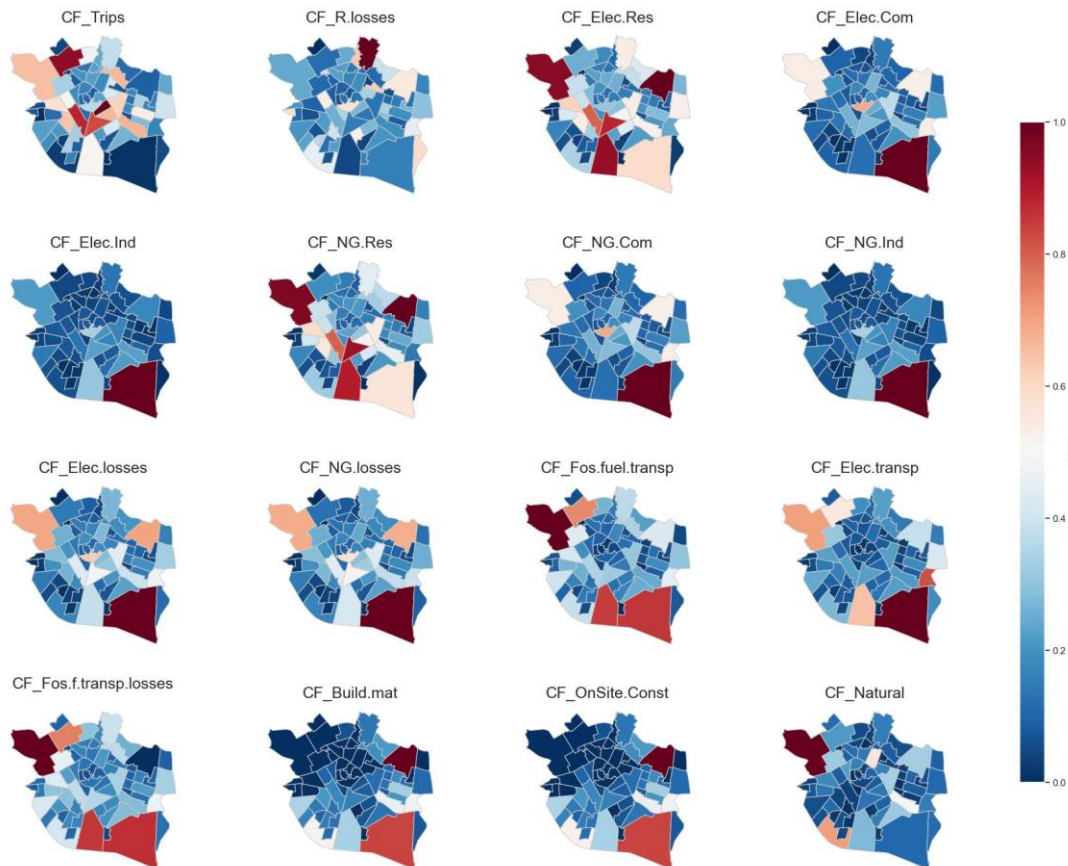
Para complementar la explicación del análisis espacial, la Figura 3.8 muestra de nuevo los mapas coropléticos para las variables que han resultado estadísticamente significativas según el índice C de Geary.

Figura 3.7.- Clasificación de los clústeres de la I_i de Anselin para los indicadores individuales estadísticamente significativos



En primer lugar, el indicador “*Trips*” cuenta con un grupo de barrios en la zona centro formado por Mont-Olivet, Russafa, Gran Vía, El Pla del Remei y Sant Francesc con clasificación HH, en contraposición a los barrios de Les Tendetes, El Carme y Morvedre que se clasifican como LL. Este comportamiento refleja que los habitantes de los primeros cinco barrios y sus vecinos realizan muchos viajes de larga distancia, lo que incrementa su huella de carbono; mientras que la población del segundo grupo de barrios y sus vecinos realizan pocos viajes, de ahí los valores bajos para este indicador. Es importante destacar que, tal y como refleja el mapa de calor para la variable “*Trips*” en la Figura 3.8, los valores tan bajos que presentan barrios como La Punta o Camí Real no han resultado estadísticamente significativos según el índice I_i de Anselin, lo que significa que sus buenos resultados no se relacionan con el entorno en el que se encuentran. Además, se identifica el barrio de Benicalap como uno de los barrios con puntuaciones más elevadas, lo que le convierte en un valor anómalo HL según la I_i de Anselin, pues sus vecinos presentan valores bajos para este indicador. Este barrio puede considerarse un outlier espacial, ya su valor difiere significativamente de sus vecinos.

Figura 3.8.- Mapas coropléticos de las variables estadísticamente significativas según la C de Geary



Fijando la atención en el indicador “*Railway Losses*”, los barrios de Benicalap, Marxalenes, Morvedre y El Carme se clasifican como LL. A su lado se encuentra el barrio de El Calvari, que se clasifica como HL por presentar valores elevados para esta variable en contraposición con los de su alrededor. También se observa un único barrio con clasificación HH, Camí de Vera, que se encuentra pegado a un barrio con valores bajos para esta variable, Benimaclet, por lo que se ve clasificado como LH. Existen otros barrios que presentan una huella de carbono elevada para este indicador, como Sant Llorenç (véase Figura 3.8). No obstante, no han resultado estadísticamente significativos, por lo que sus malos resultados no se relacionan con los barrios que le rodean.

El grupo de barrios con clasificación LL que se muestran para la variable “*Electricity Residential*” es similar al del caso anterior, añadiendo el barrio de Vara de Quart situado al oeste. En estos barrios, el consumo de electricidad en los hogares es bajo y sus buenos resultados están relacionados con los barrios vecinos. Además, cabe destacar que, para este indicador, no se observan barrios con clasificación HH, a pesar de que el mapa de

calor para esta variable refleja que los barrios de Sant Pau, Malilla o La Carrasca presentan valores elevados (véase Figura 3.8).

En el indicador “*Electricity Commercial*” destacan dos agrupaciones de barrios con valores bajos significativos (clasificación LL). Por un lado, se encuentran los barrios de Benicalap y Torrefiel y, por otro, todos los barrios situados más al oeste de Valencia, desde Nou Moles hasta Malilla. Ambos grupos de barrios se encuentran unidos por Sant Pau que, al presentar un valor elevado para esta variable en contraposición con el resto de barrios, se clasifica como HL. Finalmente, se identifican los tres barrios centrales de Valencia (Gran Vía, El Pla del Remei y La Xerea) con clasificación HH, lo que implica que este grupo tienen elevados consumos de electricidad en los establecimientos comerciales. A pesar de que el barrio de La Punta no ha resultado estadísticamente significativo, la Figura 3.8 muestra que presenta una elevada huella de carbono para este indicador, pero su resultado no se relaciona espacialmente con los resultados de sus vecinos.

Todos los resultados anteriormente descritos pueden extrapolarse al indicador “*NG Commercial*”, puesto que se observa un comportamiento idéntico en todos los barrios. Esto significa que el consumo de electricidad y de gas natural en los establecimientos comerciales de cada barrio es proporcionalmente equivalente.

Al estudiar el comportamiento de la variable “*Electricity Industrial*”, el barrio de La Punta, así como el de Ciutat de les Arts i les Ciències, resultan estadísticamente significativos con clasificación HH, lo que indica que sus elevados consumos de electricidad en el ámbito industrial se relacionan con el entorno en el que se encuentran. Estos barrios están rodeados por dos regiones con clasificación LH, Natzaret y Fonteta Sant Lluís, los cuales se identifican como anómalos por presentar valores bajos. Para este indicador, destacan también los barrios situados al norte de Valencia, desde Ciutat Fallera hasta Benimaclet, con clasificación LL. Este grupo también está rodeado por varios barrios anómalos con clasificación HL, como Sant Pau, Sant Llorenç o La Carrasca.

De manera similar a lo que ocurre con las dos variables referidas a los establecimientos comerciales, la información de los indicadores que miden los consumos de electricidad y gas natural en la industria es la misma. En otras palabras, todo lo descrito para el indicador "Electricity Industrial" se aplica a "NG Industrial" y, aunque los valores

absolutos para ambas variables pueden diferir, las proporciones para cada barrio se mantienen constantes.

En cuanto a la variable “*NG Residential*”, los barrios del norte cuentan con una clasificación LL, concretamente Benicalap, Torrefiel, Tormos, Morvedre, Trinitat y El Carme. Este grupo presenta bajos consumos de gas natural en los hogares, en contraposición con La Petxina y Mont-Olivet, los cuales se clasifican como HH. A pesar de que barrios como Sant Pau o La Carrasca tienen una elevada huella de carbono para este indicador, tal y como se observa en la Figura 3.8, sus valores no se relacionan con su situación geográfica, y por ello no han resultado estadísticamente significativos.

En los indicadores “*Electricity Losses*” y “*NG Losses*” se observa un comportamiento similar, con dos grupos de barrios con clasificación LL, concretamente la zona norte de Benicalap, Torrefiel, Tormos y Marxalenes y la zona oeste, desde Vara de Quart hasta Camí Real. Estos grupos están rodeados por dos barrios con clasificación HL, Sant Pau y Malilla. Finalmente, los dos barrios del centro, Gran Vía y El Pla del Remei resultan significativos por su clasificación HH para ambos indicadores, mientras que para “*NG Losses*” se observa también el barrio de Ciutat de les Arts i les Ciències como significativo. Esto indica que las elevadas cantidades de pérdidas de energía por consumo de electricidad y gas natural en estos barrios se relacionan con el entorno en el que se encuentran los mismos.

La información mostrada por el indicador “*Fossil Fuel Transport*” se encuentra muy relacionada con “*Fossil Fuel Transport Losses*”. En concreto, en ambas variables se observa un clúster central con clasificación LL formado por los barrios de Morvedre, Trinitat, El Carme, La Seu, La Xerea, Sant Francesc, El Pilar, El Mercat y El Botànic. Estos barrios y sus vecinos presentan una baja huella de carbono en el consumo de combustibles fósiles por el transporte, información que es coherente con la que se muestra en los mapas coropléticos de la Figura 3.8. Además, para las pérdidas generadas por este consumo, se reconoce otra región LL formada por los barrios de La Carrasca, L’Amistat, Ciutat Jardí (también significativo en el otro indicador), Beteró y La Malva-rosa. Para “*Fossil Fuel Transport*”, el barrio de La Carrasca es un outlier espacial, puesto que cuenta con un valor significativamente más elevado que sus vecinos (clasificación HL). Finalmente, en ambos casos, Mestalla también se clasifica como HL y conecta las agrupaciones LL. Para ninguno de las dos variables se observan clústeres significativamente elevados con clasificación HH.

Para la variable “*Electric Transport*” sucede algo similar a los dos casos anteriores, puesto que la zona central de Valencia forma una agrupación de valores bajos (LL) en el consumo de electricidad por los vehículos eléctricos. En conjunto, se puede destacar que los barrios céntricos de la ciudad hacen poco uso de vehículos y medios de transporte privados. En este indicador sí se observa un clúster con clasificación HH formado por los barrios de La Punta y Natzaret, cuya población realiza numerosos desplazamientos mediante vehículos eléctricos. Asimismo, se identifican los barrios de Vara de Quart, Arrancapins y Mestalla como outliers espaciales, pues tienen valores anormalmente más elevados que sus vecinos (clasificación HL). Finalmente, cabe destacar que aunque en el mapa coroplético se observan otros barrios con valores elevados para este indicador (como El Grau o Sant Pau), no han resultado estadísticamente significativos, lo que indica que sus consumos elevados no están relacionados con el entorno en el que se sitúan.

En los indicadores “*Building Material*” y “*On Site Construction*” se observa un comportamiento idéntico. La zona noroeste de Valencia formada por barrios desde Sant Pau hasta Exposició y El Pla de Remei se clasifican como LL. Esta información muestra de forma clara lo que se observa en la Figura 3.8, pues para todos estos barrios su huella de carbono relacionada con el proceso de construcción y los productos empleados para ello es muy baja. En contraposición, los barrios de La Punta y Ciutat de les Arts i les Ciències situados al sureste, así como La Creu del Grau, L’Hort de Senabre y Ciutat Jardí, se encuentran clasificados como HH. Estos cinco barrios cuentan con procesos de construcción avanzados y sus puntuaciones elevadas se relacionan con su disposición geográfica.

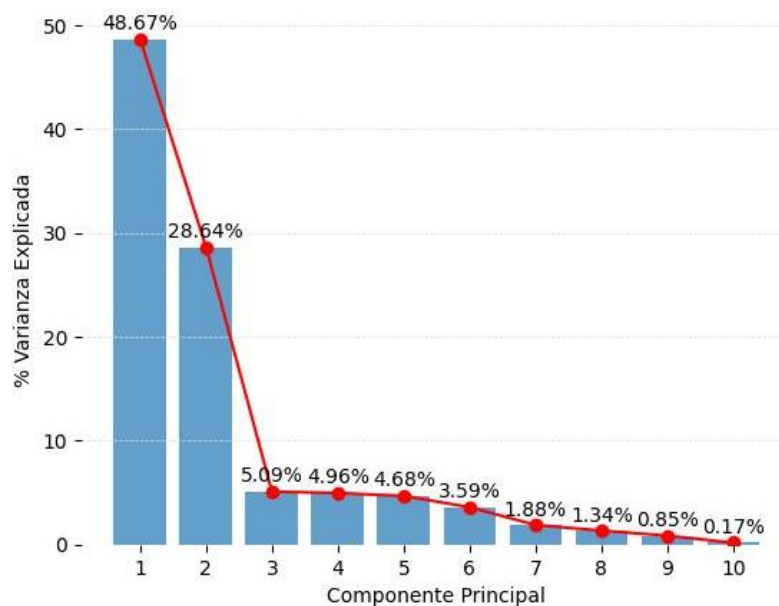
Finalmente, el indicador “*Natural Sequestration*” que recoge la absorción de carbono por las áreas verdes disponibles en cada barrio no muestra ninguna región con clasificación HH. Esto se debe a que, como se observa en el mapa coroplético de la Figura 3.8, únicamente el barrio de Sant Pau toma valores elevados para esta variable, pero su medida positiva no se relaciona con el entorno en el que se encuentra. Sí se puede destacar un grupo de barrios con clasificación LL formado por Safranar, La Raiosa, Arrancapins, La Roqueta y Sant Francesc. Estos barrios y sus vecinos presentan valores bajos para este indicador, lo que indica que cuentan con pocas zonas verdes que puedan reducir la huella de carbono. Además, se observan dos outliers espaciales con clasificación HL (Russafa y La Carrasca) que, a pesar de no presentar valores muy elevados, destacan frente a sus vecinos por tener un mejor resultado.

3.1.2. Análisis de Componentes Principales

Tras comprender la estructura espacial subyacente en los datos, se ha realizado un Análisis de Componentes Principales para los 24 indicadores individuales con el fin de comprender, desde una perspectiva multivariante, el comportamiento de los mismos y sus relaciones.

Una vez realizado el escalado de los datos a media cero y varianza unitaria, con una muestra de $n = 70$ observaciones (barrios) y $p = 24$ variables, en la Figura 3.9 se muestra el Scree Plot del PCA obtenido. Con la información mostrada, se puede concluir que la primera componente explica un porcentaje muy superior al resto de componentes, en concreto un 48,67%. La componente número seis explica un 3,59% de la variación contenida en los datos. A partir de la componente siete, se puede observar una caída más pronunciada en el porcentaje explicado. Por este motivo, se deciden mantener seis componentes principales que explican el 95,63% de la variación contenida en los datos.

Figura 3.9.- Scree Plot del PCA para los indicadores individuales



Análisis de loadings

Tras retener la información de las seis primeras componentes principales, la Tabla 3.3 muestra los *loadings* o pesos de cada uno de los indicadores sobre cada componente. Con esta información, es posible identificar qué indicadores cuentan con mayor relevancia en la conformación de la estructura latente de los datos. Esta información se puede visualizar

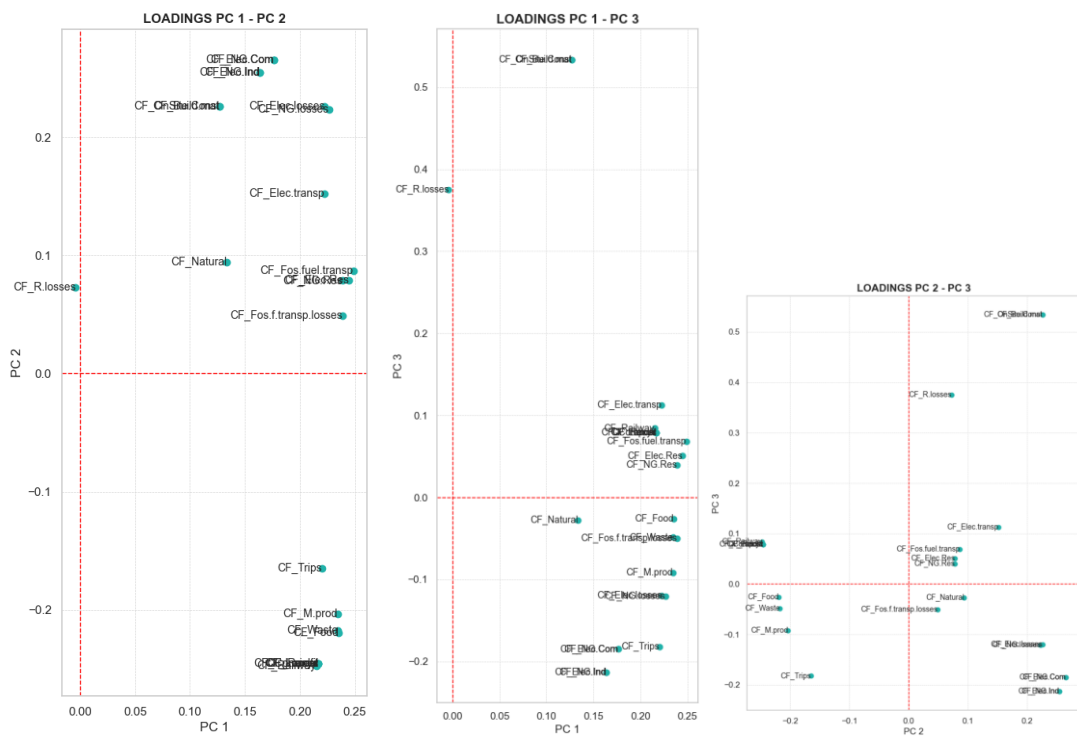
en la Figura 3.10, donde se ven representados los gráficos de loadings para las tres primeras componentes principales por pares.

Tabla 3.3.- Loadings de las seis primeras componentes principales del PCA para los indicadores individuales

Indicador	PC 1	PC 2	PC 3	PC 4	PC 5	PC 6
Food	0,235586	-0,21948	-0,026	-0,08843	-0,02356	-0,03128
M.prod	0,234208	-0,20324	-0,09203	-0,14448	-0,07707	-0,06289
Trips	0,220069	-0,16494	-0,18235	-0,21588	-0,15876	-0,11757
Waste	0,234908	-0,21727	-0,04875	-0,1101	-0,03434	-0,03321
Railway	0,215027	-0,24729	0,084159	-0,01795	0,108121	0,068719
Railway.Losses	-0,00405	0,072826	0,375156	-0,26274	-0,4859	0,703511
Elec.Resid	0,244935	0,07877	0,050503	-0,03595	-0,31333	-0,25289
Elec.Comm	0,176396	0,265443	-0,18505	-0,14736	-0,01697	0,124592
Elec.Indust	0,163564	0,254606	-0,21318	-0,12706	0,282809	0,079086
NG.Resid	0,238633	0,078544	0,0396	-0,04031	-0,32426	-0,30562
NG.Comm	0,176396	0,265443	-0,18505	-0,14736	-0,01697	0,124592
NG.Indust	0,163564	0,254606	-0,21318	-0,12706	0,282809	0,079086
Elec.Losses	0,222653	0,226288	-0,11995	-0,12198	-0,1146	-0,00505
NG.Losses	0,226506	0,223368	-0,12084	-0,12066	-0,10715	-0,04554
Fos.Fuel.Transp	0,248731	0,086688	0,068473	0,372822	0,047241	0,134955
Elec.Transp	0,222385	0,152317	0,112581	0,287709	0,183207	0,23111
Fos.F.Trans.Losses	0,238809	0,048909	-0,05023	0,420008	0,106793	0,228478
Build.Mat	0,126826	0,226285	0,533626	-0,10625	0,175734	-0,23506
OnsSite.Const	0,126826	0,226285	0,533626	-0,10625	0,175734	-0,23506
Compost	0,216715	-0,2456	0,078682	-0,00879	0,106502	0,070223
Recycling	0,216715	-0,2456	0,078682	-0,00879	0,106502	0,070223
Landfilling	0,216715	-0,2456	0,078682	-0,00879	0,106502	0,070223
Incineration	0,216715	-0,2456	0,078682	-0,00879	0,106502	0,070223
Natural.Seq	0,133575	0,093944	-0,02751	0,560427	-0,42015	-0,15304

De la información mostrada en los gráficos de loadings se puede concluir que la primera componente (PC 1) recoge la mayor parte de la información, como las variables “*Fossil Fuel Transport*”, “*Electricity Residential*” o “*NG Residential*” destacables por tener un mayor peso. Estas variables, así como las que se encuentran próximas a estas, se sitúan en la parte derecha y próximas al eje horizontal en los gráficos de loadings para PC 1 – PC 2 y PC 1 – PC 3. Para todas las variables se observa una asociación lineal positiva entre la primera variable latente y los indicadores de huella de carbono.

Figura 3.10.- Gráficos de loadings para las tres primeras componentes del PCA



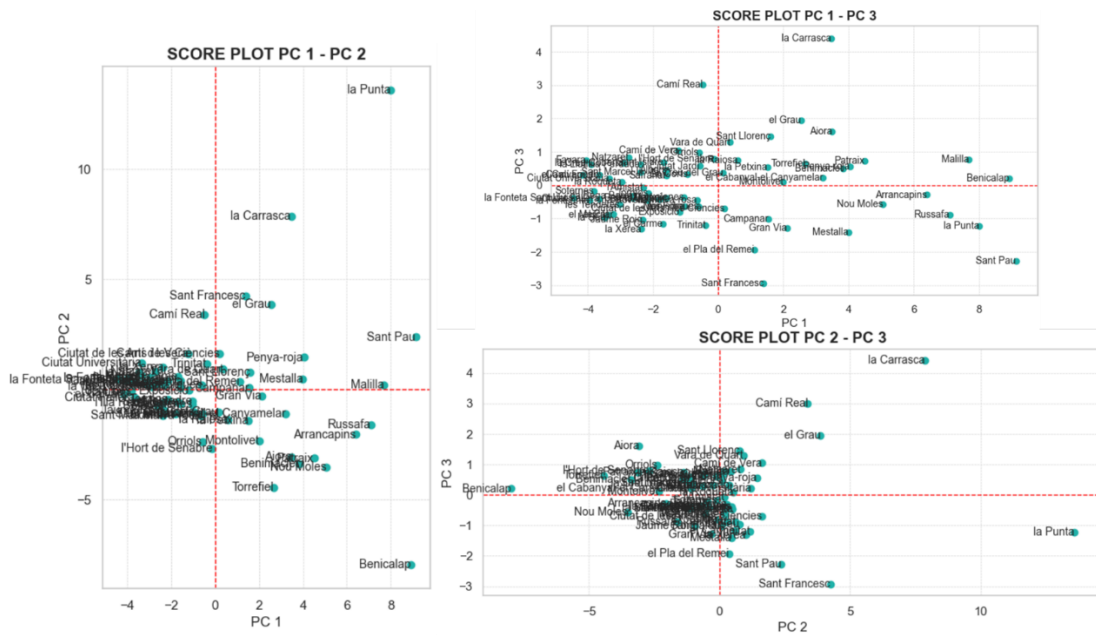
Del mismo modo, la segunda componente principal (PC 2) explica la información de las variables relacionadas con el consumo de electricidad y gas natural en comercios e industria (“*Electricity Commercial*”, “*Electricity Industrial*”, “*NG Commercial*” y “*NG Industrial*”), entre otras. De nuevo, existe una relación lineal positiva entre estas variables y la segunda variable latente. Sin embargo, esta variable latente también recoge la información de variables como “*Compost*”, “*Recycling*”, “*Landfilling*” o “*Incineration*”, las cuales presentan una relación negativa con la segunda variable latente.

Finalmente, la tercera componente principal (PC 3) explica un menor porcentaje de variación que las dos anteriores, destacándose las dos variables relacionadas con la construcción (“*Bulding Material*” y “*On Site Construction*”), así como “*Railway Losses*”. Las tres componentes principales siguientes van recogiendo cada vez menos información.

Análisis de scores

Una vez se ha observado la estructura latente de las seis componentes principales retenidas, a continuación se relaciona la información obtenida de los loadings con los valores obtenidos de los scores, con el fin de analizar las tendencias de los barrios en cada componente. Del mismo modo que antes, en la Figura 3.11 se presentan los gráficos de scores para las tres primeras componentes por pares.

Figura 3.11.- Gráficos de scores para las tres primeras componentes del PCA



Centrando la atención en los barrios con puntuaciones más elevadas en la PC 1, destacan barrios como Sant Pau, Malilla o Russafa. Estos barrios tienden a contar con valores más elevados en los indicadores con loadings más influyentes para la PC 1, como el consumo de electricidad y gas natural residencial. En el otro extremo se encuentra barrios como Soternes, Ciutat Universitària, Favara o Fonteta Sant Lluís, los cuales presentan una baja huella de carbono para los indicadores anteriormente mencionados.

Por otra parte, los barrios de La Carrasca, Sant Francesc o Camí Real presentan elevados scores para la PC 2, cuyo comportamiento refleja que tienen elevados consumos en electricidad y gas natural en comercios e industria. Asimismo, destaca el barrio de La Punta con scores elevados tanto en la PC 1 como en la PC 2 (esquina superior derecha), lo que señala una elevada huella de carbono en los indicadores anteriormente comentados. Otro bario a destacar es Benicalap (esquina inferior derecha), el cual presenta elevados valores positivos para la PC 1 y negativos para la PC 2. Este comportamiento parece indicar que Benicalap tiene una huella de carbono elevada para las variables “Compost”, “Recycling”, “Landfilling” e “Incineration”, puesto que estos indicadores tenían una relación negativa con la segunda variable latente.

3.2. Indicadores agrupados de huella de carbono

Tal y como se desarrolló en el capítulo 3, el GPC dictamina que los indicadores de huella de carbono pueden agruparse en tres alcances en función de dónde y qué genera

esas emisiones. Consecuentemente, los indicadores individuales se han agrupado en esos tres alcances, obteniéndose cinco nuevas variables a analizar: *S1*, *S2*, *S3*, *S1S2* y *Total*.

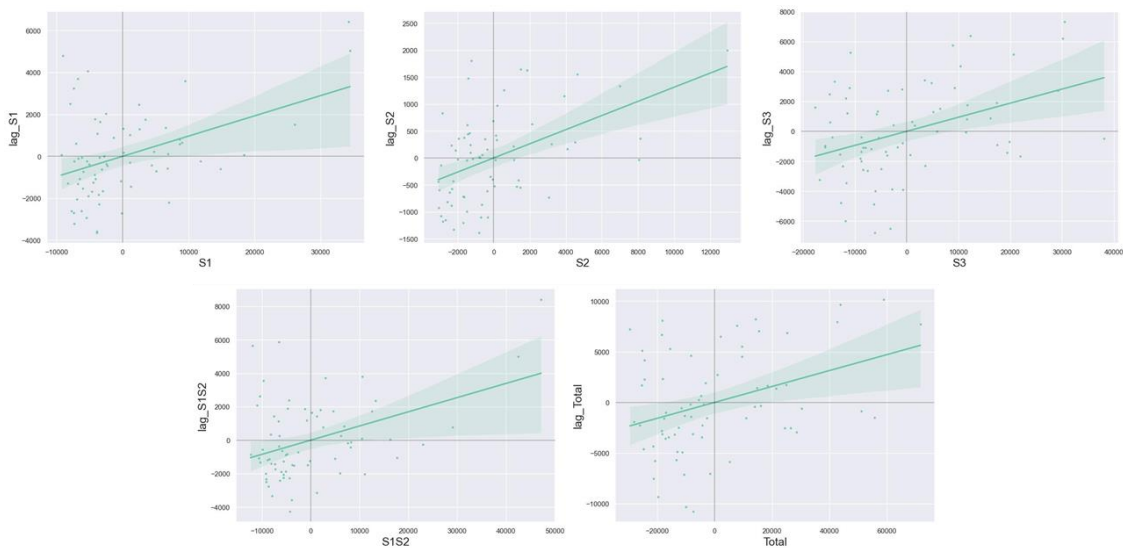
3.2.1. Análisis estadístico espacial

Tomando como referencia para el estudio de la autocorrelación espacial la matriz de pesos espaciales basados en kernel gaussiano calculada al inicio de este capítulo, se procede a realizar un análisis análogo al llevado a cabo para los indicadores individuales.

Autocorrelación espacial global

En primer lugar, con el fin de analizar la relación entre cada variable y su retardo espacial, la Figura 3.12 expone los gráficos de Moran para cada uno de los cinco indicadores agrupados. Con la información mostrada, se concluye que existe una autocorrelación espacial positiva entre cada indicador y su retardo espacial.

Figura 3.12.- Gráfico de Moran para los 5 indicadores agrupados



Una vez se ha observado que, a priori, existe una autocorrelación espacial positiva, la información mostrada en la Tabla 3.4 permite cuantificar esta relación y, gracias a los índices *I de Moran* y *C de Geary*, analizar si esta relación espacial es estadísticamente significativa o no.

Fijando la atención en los p-valores asociados a cada indicador, según el índice *I de Moran* no existe una tendencia general de agrupamiento en la Ciutat Central para ninguno de los cinco indicadores. En consecuencia, la distribución espacial global de los indicadores agrupados es aleatoria.

Tabla 3.4.- Estadísticos I de Moran y C de Geary para los indicadores agrupados

Indicador	I de Moran	p-valor (I)	C de Geary	p-valor (C)
S1	0,09652	0,481	0,74165	0,023
S2	0,13134	0,231	0,72670	0,009
S3	0,09411	0,478	0,86017	0,266
S1S2	0,08475	0,347	0,75153	0,020
Total	0,07864	0,297	0,82291	0,111

Si se analizan los p-valores obtenidos para la *C de Geary*, considerando un nivel de significación del 90% se observa significancia estadística para los indicadores S1, S2 y la suma de ambos. Por tanto, puede decirse que existen agrupamientos locales significativos de valores similares en regiones específicas para las emisiones de GEI generadas por fuentes situadas dentro de los límites urbanos (alcances 1 y 2), pero no para aquellas que se generan fuera de estos límites (alcance 3).

En conjunto, a pesar de que para ninguno de los alcances existe un patrón global claro de clustering donde se agrupen áreas con valores altos o bajos, para las emisiones de GEI generadas por fuentes situadas dentro de los límites de cada barrio puede observarse una variación estadísticamente significativa en los valores de los barrios vecinos. Para estudiar en qué regiones específicas ocurre esto, se deben estudiar la autocorrelación espacial local.

Autocorrelación espacial local

El estadístico I_i de Anselin permite estudiar, para las emisiones de alcance 1 y 2, en qué grupos de barrios se agrupan tipos específicos de valores y si existen outliers espaciales en la ciudad. Previamente, en los gráficos KDE mostrados en la Figura 3.13 se visualiza un ligero sesgo a la izquierda en la distribución del estadístico, lo que deriva en la existencia de autocorrelación espacial positiva que ya adelantaban los gráficos de Moran.

Para cuantificar estadísticamente esta información y visualizar las regiones significativas para un nivel de significación del 90%, la Figura 3.14 muestra el mapa de Valencia para cada indicador según los valores obtenidos de a I_i de Anselin.

Figura 3.13.- Gráfico KDE del estadístico I_i de Anselin para los indicadores agrupados

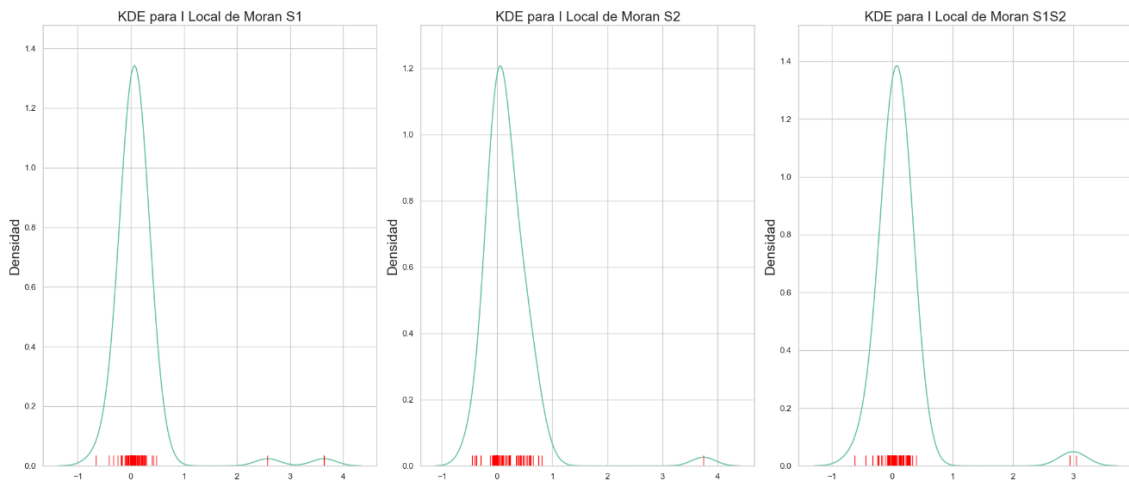
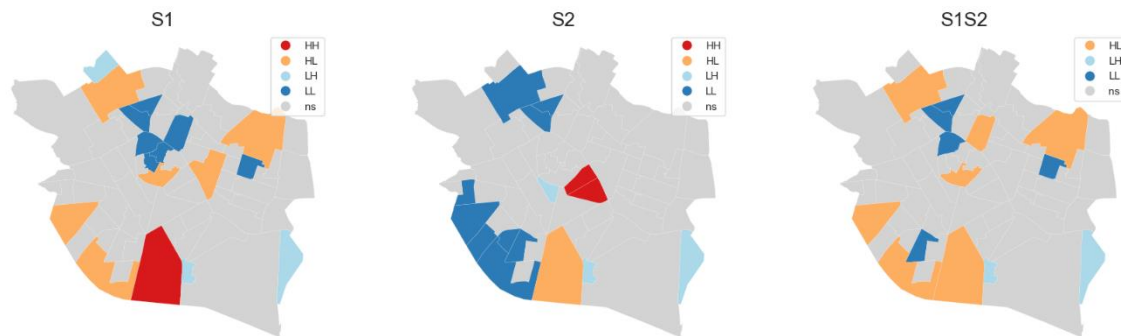
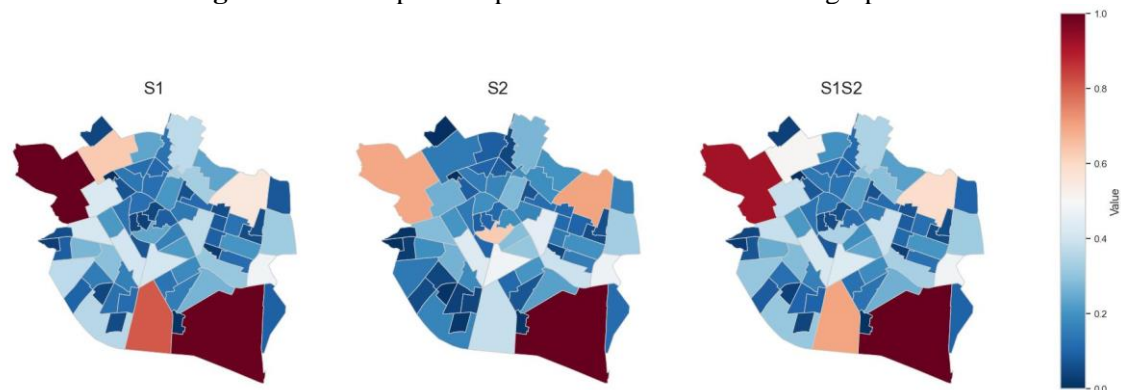


Figura 3.14.- Clasificación de los clústeres de la I_i de Anselin para los indicadores agrupados



Asimismo, tal y como se realizó para los indicadores individuales, con el fin de facilitar la comprensión de los resultados y realizar una explicación más completa, se han elaborado mapas coropléticos para los indicadores agrupados estadísticamente significativos según la C de Geary (véase Figura 3.15).

Figura 3.15.- Mapas coropléticos de los indicadores agrupados



Analizando conjuntamente toda la información, puede observarse que el barrio de Malilla presenta una clasificación HH para el indicador S1, lo que indica que tanto este

como sus vecinos tienen valores altos para esta variable. Sin embargo, en S2 se considera un outlier espacial, ya que aunque presenta valores elevados para el alcance 2, sus vecinos tienen valores bajos (clasificación HL). Este comportamiento anómalo se mantiene para la suma de ambos indicadores.

El único clúster con clasificación HH en S2 lo forman los barrios centrales de Gran Vía y El Pla del Remei, lo que puede considerarse que tanto ellos como sus vecinos consumen grandes cantidades de energía. Para la suma de los alcances 1 y 2 no se observa ningún barrio estadísticamente significativo que comparta con sus vecinos valores altos.

Los barrios situados al suroeste de Valencia (La Font Santa, Vara de Quiart, Sant Isidre, Favara, L'Hort de Senabre y Camí Real) forman un clúster con clasificación LL para S2. No obstante, Vara de Quart y Camí Real representan outliers espaciales en S1, puesto que ellos mismos cuentan con valores altos y sus vecinos con valores bajos (clasificación HL). Si se recopila toda la información en el indicador S1S2, L'Hort de Senabre mantiene su comportamiento LL derivado de S2, mientras que Vara de Quart y Camí Real heredan su comportamiento de S1, haciendo que pesen más sus altos valores en S1 que sus bajos valores en S2.

A partir de lo mencionado anteriormente, se puede concluir que, según los indicadores agrupados de huella de carbono, no existe una relación espacial entre los barrios de la ciudad de Valencia. En términos generales, no se observan agrupaciones de barrios relacionados por su cercanía con una huella de carbono significativamente mayor o menor de manera agrupada. Para observar diferencias significativas a nivel espacial entre barrios se necesita un mayor nivel de detalle en los indicadores de huella de carbono, desagregándolos en indicadores individuales.

3.2.2. Métodos de agrupación

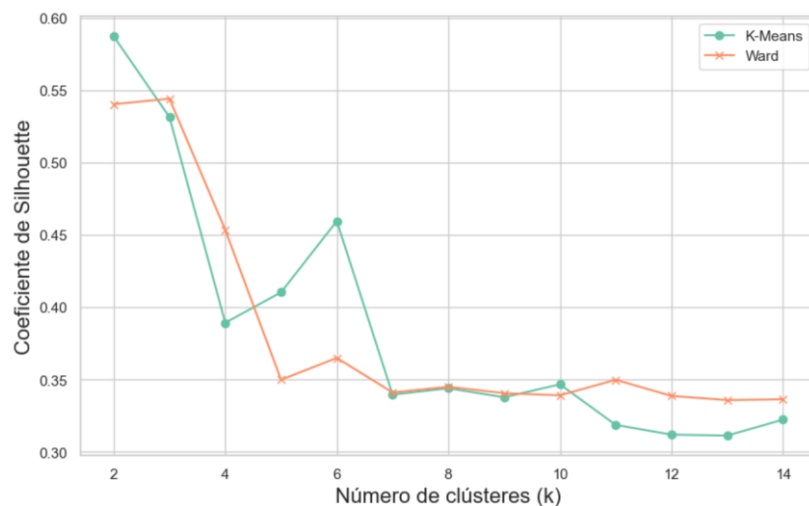
Una vez se ha comprendido a nivel univariante la estructura espacial de los indicadores agrupados, resulta de gran interés comprender su estructura a nivel multivariante. Para ello, la presente sección está dedicada a explorar los resultados obtenidos de la aplicación del análisis clúster y la regionalización, llevados a cabo con el objetivo de identificar posibles patrones ocultos en los datos y analizar si es posible agrupar los barrios en grupos con características similares.

Análisis clúster

En primer lugar, con el fin de realizar una comparativa entre diversos métodos de análisis clúster, se han aplicado dos algoritmos de agrupamiento diferentes: por un lado, el clustering no jerárquico de K-Means y, por otro lado, el clustering jerárquico aglomerativo siguiendo el método de Ward. Antes de proceder con la implementación de los resultados, se han normalizado las variables para que tengan media cero y varianza unitaria.

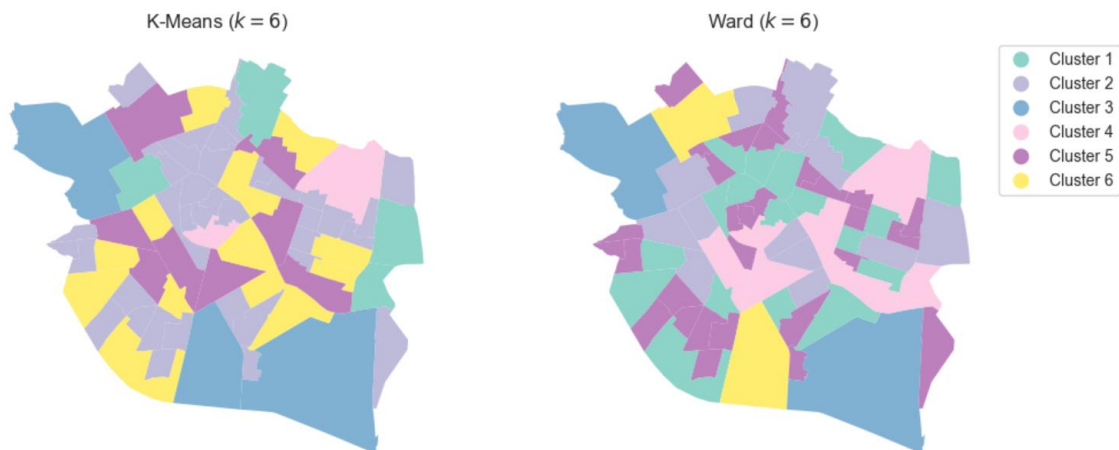
Para poder aplicar estos métodos, se debe establecer previamente el número de clústeres (k). Para ello, se ha calculado el coeficiente de Silhouette para cada valor de k , tanto para el método de K-Means como para el método de Ward. En la Figura 3.16 se muestran los resultados.

Figura 3.16.- Comparación coeficiente de Silhouette para clustering K-Means y Ward



Observando las gráficas, el número óptimo de clústeres para K-Means es $k = 2$ y para Ward es $k = 3$. No obstante, se considera una agrupación insuficiente y se decide seleccionar un número de clústeres algo superior e igual a $k = 6$ en ambos casos, pues es donde se observa un crecimiento en el coeficiente de Silhouette.

La Figura 3.17 refleja en un mapa los seis clústeres obtenidos según el método K-Means y el método de Ward. Comparando los resultados, se observa que el método K-Means agrupa en el clúster número 1 los barrios de Sant Llorenç, Campanar, El Cabanyal-El Canyamelar y El Grau; mientras que para el método de Ward los tres primeros barrios se agrupan en un mismo clúster (número 2) y El Grau se agrupa en el clúster 4.

Figura 3.17.- Clústeres obtenidos según el metodo K-Means (izquierda) y Ward (derecha)

Asimismo, el método de Ward recoge los barrios de Benicalap y Malilla en el clúster número 6 y Sant Pau y La Punta en el clúster 3. K-Means, por su parte, también agrupa estos dos últimos barrios en el clúster 3, pero también incluye a Malilla.

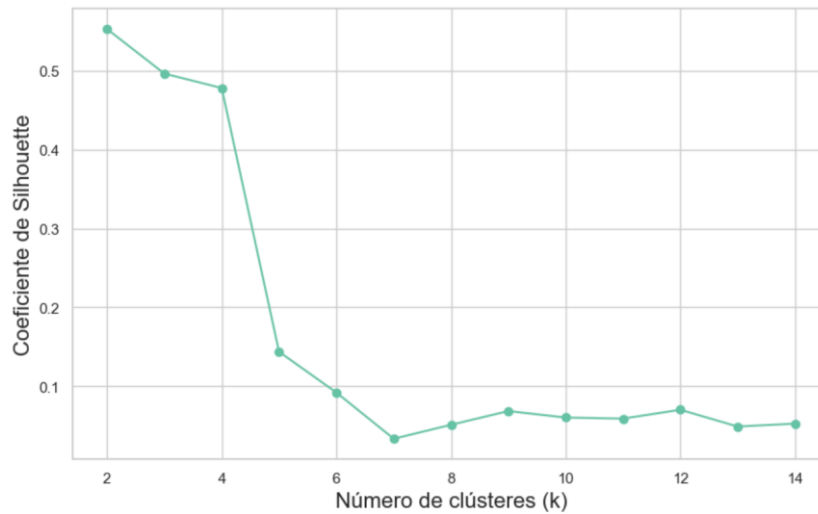
Otra de las diferencias a destacar es que K-Means agrupa en el clúster 4 los barrios de Sant Francesc y La Carrasca, mientras que el método de Ward amplía este clúster recopilando un total de 7 barrios.

El clúster más numeroso para K-Means es el 2, con un total de 38 barrios, mientras que para el método de Ward es el número 5, con 29 barrios.

Regionalización

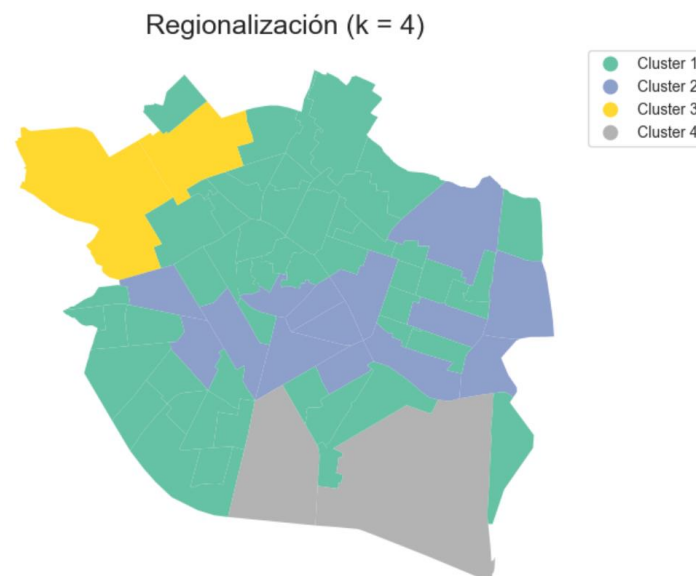
La regionalización incorpora restricciones espaciales para generar clústeres de regiones completamente conectadas. Para realizar este análisis se hace uso de la misma matriz de pesos espaciales calculada para el propio análisis estadístico espacial, cuyos pesos están basados en una función kernel gaussiana con $k = 10$.

De manera similar a lo realizado para los dos métodos de clustering anteriores, el coeficiente de Silhouette para el análisis de regionalización se ve representado en la Figura 3.18. Con los resultados mostrados, el número óptimo de clústeres es $k = 2$. No obstante, al considerarse un número insuficiente y observarse una caída pronunciada a partir de 4, se ha optado por seleccionar $k = 4$.

Figura 3.18.- Coeficiente de Silhouette para el análisis de regionalización

En la Figura 3.19 se pueden visualizar los barrios agrupados en cada uno de los cinco clústeres obtenidos mediante el análisis de regionalización. Al introducir la restricción espacial, la regionalización ha intentado formar grupos de barrios conectados entre sí.

En concreto, los clústeres 3 y 4 han conseguido formar regiones compactas de barrios; sin embargo, para los clústeres 1 y 2 este análisis no ha dado como resultado distribuciones de barrios concentrados con grupos completamente conectados. Por tanto, los resultados revelan que, en conjunto, no es posible formar clústeres de barrios próximos entre sí con características similares.

Figura 3.19.- Clústeres obtenidos del análisis de regionalización

Evaluación de los métodos de agrupación

Con el fin de contrastar cuantitativamente y de manera objetiva los resultados obtenidos de las tres técnicas de clustering empleadas, se procede a calcular una serie de métricas. En concreto, a continuación se estudian métricas que evalúan la similitud entre los métodos, la coherencia geográfica y la bondad de ajuste de cada técnica.

- 1. Similitud entre métodos.** Para comparar si los clústeres formados por cada uno de los métodos son similares, se hace uso del *Adjusted Mutual Information* o AMI (Vinh et al., 2009). Esta medida toma el valor 1 cuando los agrupamientos formados son idénticos, por lo que cuanto más se aproxime a 1, más semejantes serán dos métodos. Como se observa en la Tabla 3.5, los agrupamientos sin restricciones espaciales son los más similares entre sí, mientras que la regionalización es más disímil a los dos métodos anteriores, tal y como cabría esperar. No obstante, se observan valores muy parejos, lo que refleja que la regionalización realizada no consigue capturar todas las restricciones espaciales.

Tabla 3.5.- Valores de AMI de similitud entre los tres métodos de análisis clúster

	K-Means (k = 6)	Ward (k = 6)	Regionalización (k = 4)
K-Means (k = 6)	1	0,55093	0,44665
Ward (k = 6)	0,55093	1	0,44842

- 2. Coherencia geográfica.** La coherencia espacial se puede analizar observando si las formas de las áreas agrupadas son compactas mediante el cociente isoperimétrico o IPQ (Horn et al., 1993). Este índice compara el área de una región i con el área de un círculo c con el mismo perímetro que la región i , estableciéndose que el perímetro de i es igual a la circunferencia de c , por lo que $P_i = 2\pi r_c$. En consecuencia, el área del círculo será igual a $A_c = \pi r_c^2$. De manera intuitiva, el IPQ para cada región i se obtiene de la siguiente forma:

$$IPQ_i = \frac{A_i}{A_c} = \frac{A_i}{\pi \left(\frac{P_i}{2\pi}\right)^2} = \frac{4\pi A_i}{P_i^2}$$

El valor del cociente isoperimétrico fluctúa entre 0 y 1. Las regiones más compactas tienen un valor de IPQ próximo a 1, mientras que valores cercanos a 0 indican que se trata de regiones más alargadas o delgadas. La Tabla 3.6 recoge los valores del cociente IPQ para los tres métodos de agrupación calculados.

Tabla 3.6.- Valores de IPQ de coherencia geográfica para los clústeres obtenidos con los tres métodos de clustering

	K-Means (k = 6)	Ward (k = 6)	Regionalización (k = 4)
Clúster 1	0,16057	0,03771	0,02462
Clúster 2	0,02404	0,06234	0,04559
Clúster 3	0,21058	0,27508	0,35469
Clúster 4	0,26867	0,08515	0,50436
Clúster 5	0,07734	0,02612	-
Clúster 6	0,04311	0,32636	-

Pese a que los clústeres 1 y 2 violan la restricción espacial impuesta en la regionalización, la coherencia espacial de los clústeres 3 y 4 en el análisis de regionalización es mejor a la de las técnicas de clustering tradicionales, como era de esperar. Esto se debe a que la regionalización impone la restricción espacial, formando clústeres más compactos y menos dispersos. En particular, el clúster número 4 es el que mayor valor de IPQ tiene para la regionalización (0,504), pues está formado por los barrios vecinos de Malilla y La Punta. Algo similar sucede para el clúster 6 del método de Ward, que tiene un valor más elevado al resto (0,326) debido a que agrupa dos barrios que no presentan formas muy irregulares.

Sin embargo, la coherencia geográfica obtenida mediante la regionalización supone un coste en términos de bondad de ajuste.

3. Bondad de ajuste. Para medir la coherencia de las características o bondad de ajuste de las distintas agrupaciones se ha optado por el índice de Calinski Harabasz o CH (Calinski & Harabasz, 1974). Esta métrica evalúa cómo están de separados o compactos los clústeres al comparar la varianza entre grupos con la varianza dentro de los grupos. Los valores del score CH pueden variar según la estructura de los datos, pero cuanto mayor sea el valor de CH en comparación con otros métodos, mejor será la agrupación realizada por ese método. La Tabla 3.7 muestra los valores del índice CH para los tres métodos de agrupación estudiados.

Tabla 3.7.- Valores del score de CH de bondad de ajuste para los tres métodos de clustering

Score	K-Means (k = 6)	Ward (k = 6)	Regionalización (k = 4)
CH	78,361	98,479	54,983

Al analizar los resultados, se observa que la estructura de clústeres obtenida mediante el método de Ward es la que mejor se ajusta a los datos, pues toma el valor de CH toma

el valor más elevado (98,48). En consecuencia, se seleccionaría esta agrupación como la que mejor clasifica. El método K-Means toma un valor más bajo (78,36) pero cercano al anterior, lo que refleja que las técnicas de clustering sin restricciones espaciales ofrecen mejores resultados.

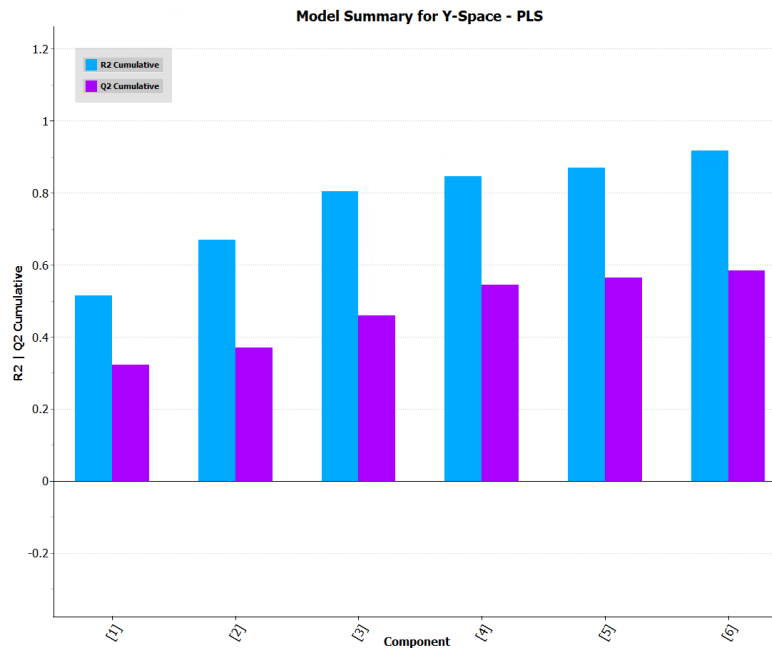
Como se comentó anteriormente, la regionalización conlleva un sacrificio en términos de bondad de ajuste debido a que, según este índice, los clústeres formados deben estar bien separados. Además, esta técnica está restringida por las restricciones espaciales que impone.

3.3. Relación entre los indicadores agrupados de huella de carbono y las variables socioeconómicas

Con el fin de comprender los comportamientos económicos y sociales de cada uno de los 70 barrios que forman la Ciutat Central de Valencia, el presente apartado está dedicado a analizar la relación existente entre los cinco indicadores agrupados que miden la huella de carbono en cada barrio, los cuales constituirán las variables Y a predecir del modelo, y las 284 variables socioeconómicas que definen sus características, que componen las variables X. Para ello, se procede a implementar la técnica de análisis multivariante Mínimos Cuadrados Parciales (PLS) con ayuda del software Aspen ProMV (versión 10.0).

Ajustando el modelo PLS de manera automática con Aspen ProMV mediante validación cruzada, en la Figura 3.20 se presentan los resultados. Inicialmente, se han extraído 6 componentes principales que explican el 91,84% de la variabilidad presente en los datos, con una bondad de predicción Q^2 del 58,61%. La componente 7 ya toma un valor negativo para la Q^2 , lo que refleja un sobreajuste del modelo. En consecuencia, se recoge el modelo con 6 componentes principales.

Cabe mencionar que la capacidad predictiva del modelo no alcanza el 60%, siendo significativamente más reducido que la bondad de ajuste o R^2 del modelo. Esto puede ser un indicativo de que existen variables incluidas en el modelo que no tienen capacidad explicativa para las variables Y incluidas, lo cual se verificará en análisis posteriores.

Figura 3.20.- R^2 y Q^2 acumuladas para cada componente principal del modelo PLS

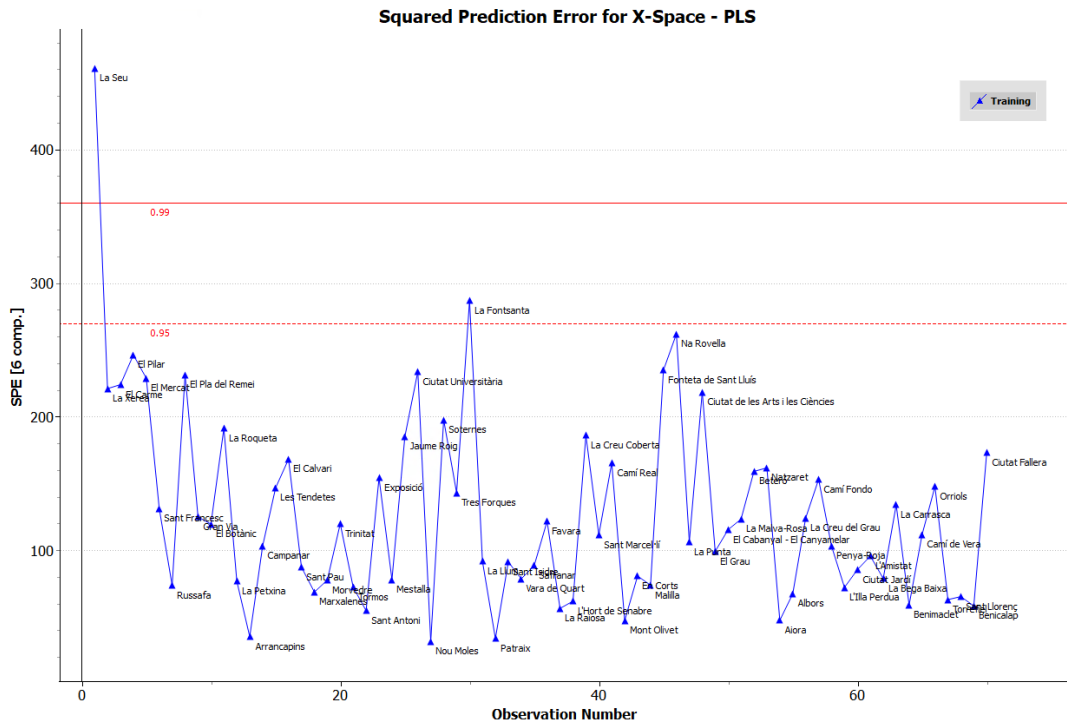
Antes de proceder a dar por válido el modelo, se debe realizar un diagnóstico de los datos. Por un lado, se lleva a cabo el análisis de los residuos con el gráfico SPE y, por otro lado, el análisis de las observaciones en el espacio de las seis componentes principales con ayuda del gráfico de T^2 de Hotelling.

3.3.1. Diagnóstico y validación de observaciones

Con el propósito de analizar la posible presencia de barrios (observaciones) que pueden considerarse anómalos en la base de datos y que afecten a los resultados del análisis, se ha realizado un diagnóstico de observaciones a partir del gráfico *Squared Prediction Error* (SPE) y el gráfico T^2 de Hotelling. Por un lado, al observar el gráfico SPE representado en la Figura 3.21, se evidencia que el barrio de La Seu sobrepasa el límite de control establecido al 99%.

Con el objetivo de identificar qué variables contribuyen a que el barrio de La Seu pueda ser una observación atípica en el modelo PLS, en la Figura 3.22 se ha extraído el gráfico de contribuciones de las variables para este barrio.

Figura 3.21.- Gráfico SPE del modelo PLS con 6 componentes principales



En el gráfico de contribuciones se observa que las variables número total de vehículos autobús, el número de personas que se dedican a trabajos relacionados con ayuda a las familias, el total de inactivos hombres, así como el número total de vehículos particulares rompen con la estructura de correlación para el barrio de La Seu en el modelo PLS. Para contemplar si, efectivamente, sucede esto, en la Figura 3.23 se muestra el gráfico de dispersión entre “Vehículos autobús” y “Turismos particulares”.

Figura 3.22.- Gráfico de contribuciones para el barrio de La Seu en el modelo PLS

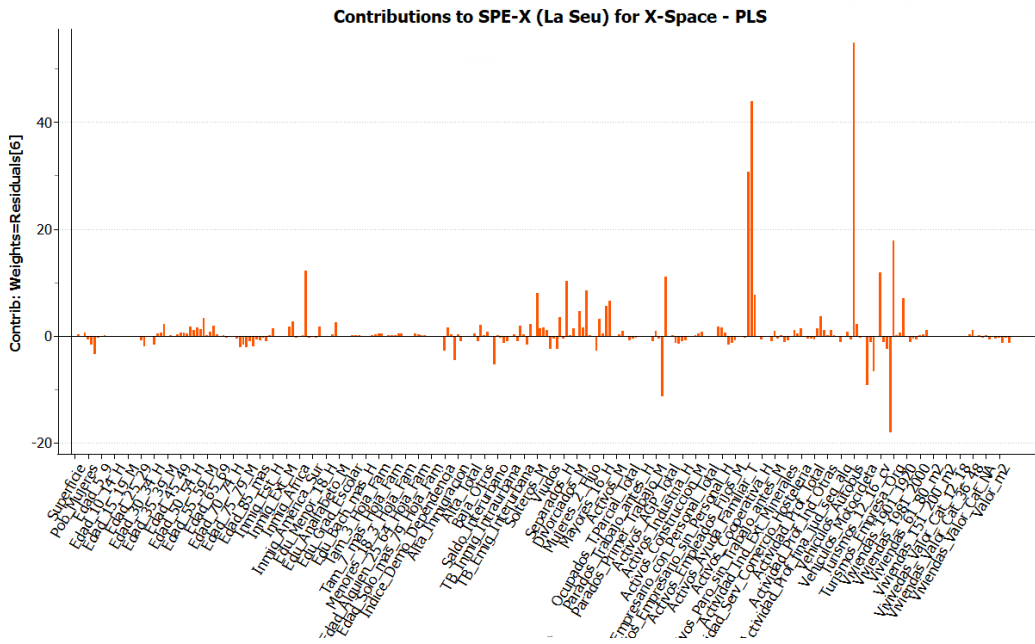
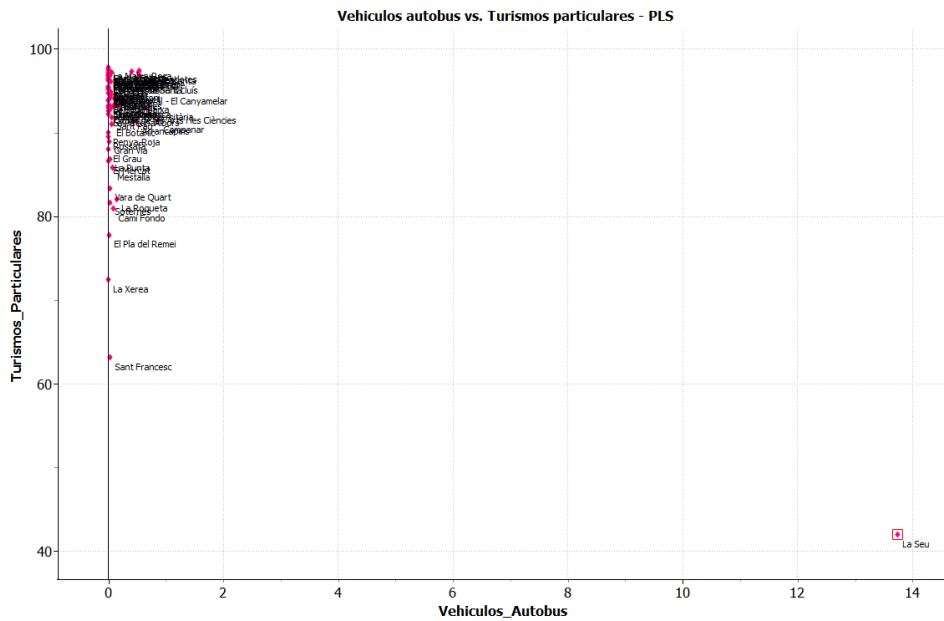


Figura 3.23.- Gráfico de dispersión entre las variables “vehículos autobús” y “turismos particulares”

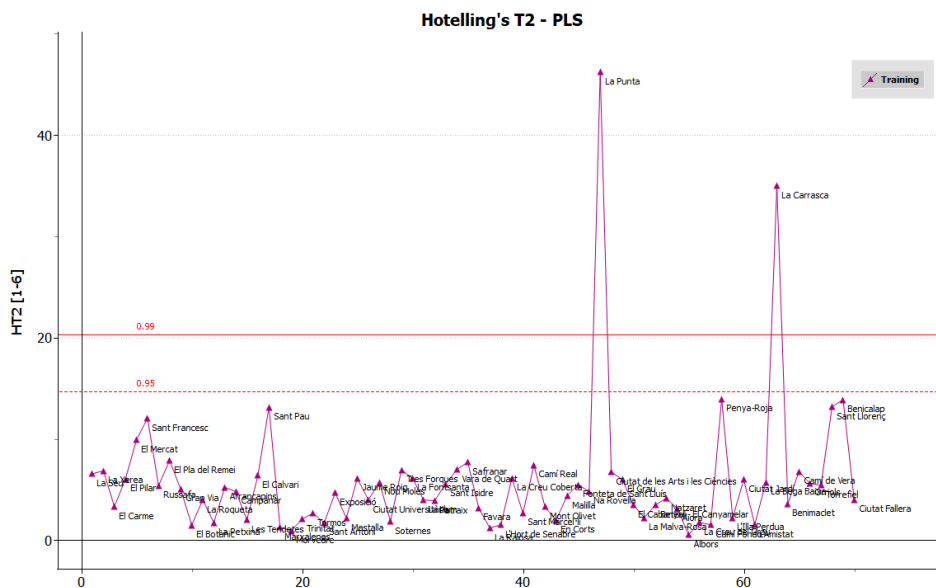


Como se exhibe, el barrio de La Seu rompe completamente la estructura de correlación entre ambas variables, pues se ha desplazado completamente de la nube de puntos del resto de barrios, situándose en la esquina inferior derecha del gráfico. Mientras que el resto de barrios tienen un número de autobuses matriculados cercano a cero o cero y un número de turismos particulares elevados, La Seu presenta una estructura completamente contraria. No obstante, el límite de control del SPE para un nivel de

confianza del 99% es 360,39 y La Seu tiene un valor del SPE de 460,92 el cual no excede dos veces el valor del límite de control. En consecuencia, se decide continuar con el análisis manteniendo esta observación en el modelo.

Por otra parte, se ha realizado el gráfico de la T^2 de Hotelling para el modelo PLS, evidenciándose en la Figura 3.24 que hay dos observaciones extremas: los barrios de La Punta y La Carrasca, los cuales presentan valores anormalmente elevados para algunas variables.

Figura 3.24.- Gráfico T^2 de Hotelling del modelo PLS con 6 componentes principales



En el caso del barrio de La Punta, dado que su valor para la T^2 de Hotelling es de 46,21, más del doble del límite de control al 99% establecido en 20,33, se ha considerado como un outlier severo y se ha eliminado del modelo.

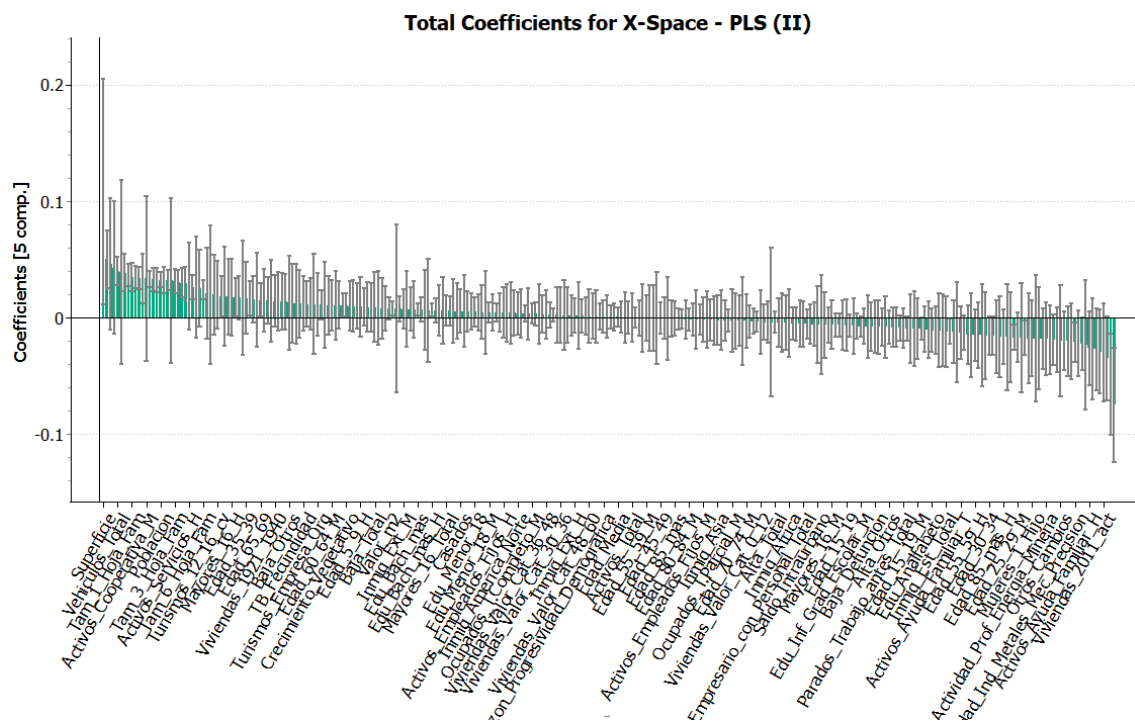
3.3.2. Validación de variables

Tras la validación de las observaciones del modelo PLS, se ha llevado a cabo un análisis de las variables incluidas en el mismo con el objetivo de excluir aquellas que no afectan a la huella de carbono total de cada barrio. Este procedimiento se ha realizado examinando los coeficientes de regresión para cada variable.

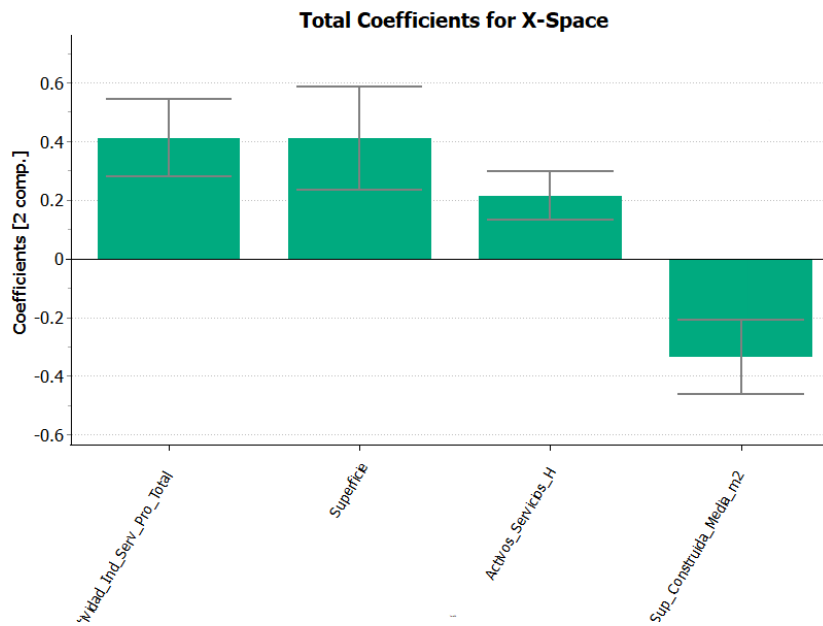
Tal y como se muestra en la Figura 3.25, existe un gran número de variables que no pueden considerarse estadísticamente significativas y que, por tanto, no influyen en la huella de carbono total existente en cada barrio. En consecuencia, mediante un procedimiento iterativo, se han ido eliminando del modelo aquellas variables

estadísticamente no significativas. Como resultado, se han mantenido en el modelo cuatro variables socioeconómicas: el número de habitantes dedicados a actividades industriales y de servicios profesionales, la superficie total de cada barrio, el número de hombres que trabajan en el sector servicios y la superficie construida media en cada barrio (véase Figura 3.26).

Figura 3.25.- Coeficientes de regresión para las variables socioeconómicas del modelo PLS



Es importante destacar que todas las variables presentan un coeficiente de regresión positivo en el modelo, a excepción de la superficie construida media en cada barrio. Esto implica que la huella de carbono total de cada barrio será mayor cuanto menor superficie de terreno haya sido construida. Por el contrario, cuanto más extenso sea un barrio, más personas trabajen en actividades industriales y de servicios profesionales, y más hombres haya empleados en el sector servicios, mayor será la huella de carbono total en ese barrio.

Figura 3.26.- Coeficientes de regresión para las variables mantenidas del modelo PLS

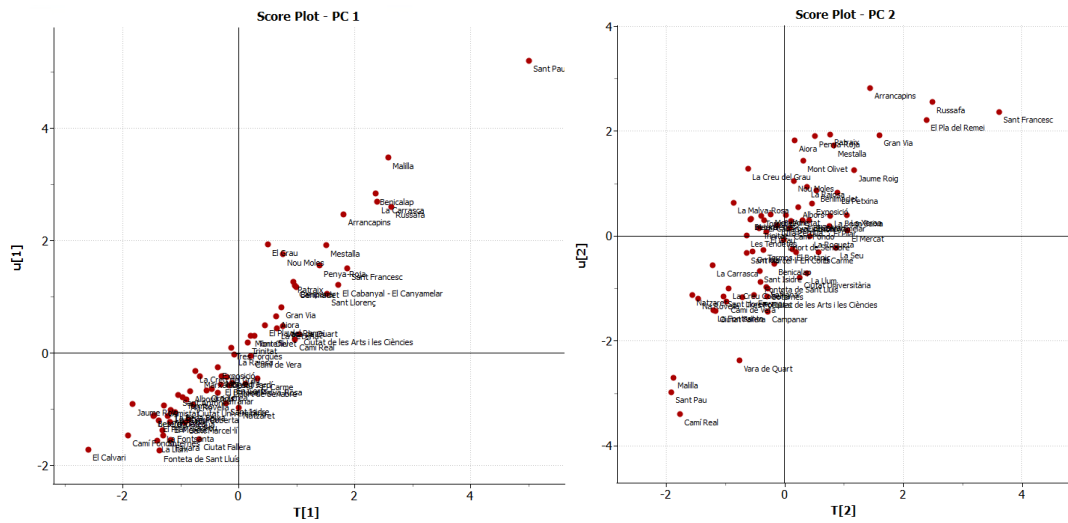
Tras haber validado por completo el modelo PLS y haber comprobado que, efectivamente, no existen barrios atípicos o extremos, se dan por válidos los resultados mostrados en la Tabla 3.8.

Tabla 3.8.- Resumen R^2 y Q^2 explicado por cada componente principal extraída del modelo PLS definitivo

Componente principal	R^2	Q^2
CP 1	0,78763	0,76363
CP 2	0,03236	0,03228
TOTAL	0,81999	0,79591

El modelo PLS definitivo ha retenido dos componentes principales que explican un 82% de la variabilidad presente en los datos. Comparando estos resultados con los iniciales, la bondad de predicción del modelo se ha incrementado hasta el 79,59%, valor que ya se asemeja mucho al de la R^2 . Además, cabe destacar que la primera componente principal es capaz de explicar, por sí sola, un 78,76% de la relación existente entre las cuatro variables socioeconómicas y los indicadores de la huella de carbono.

Figura 3.27.- Score Plot entre los scores de X e Y para las dos componentes principales

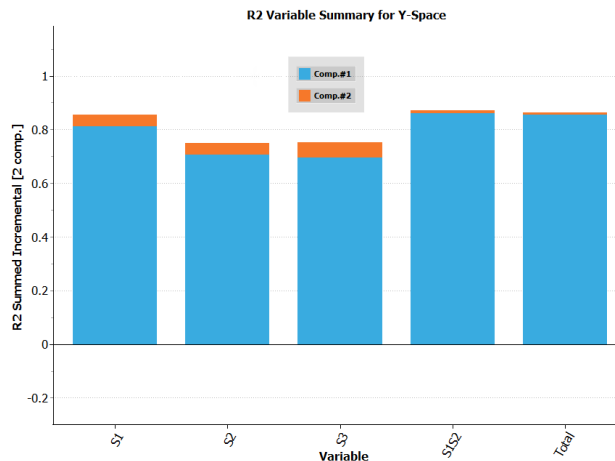


Finalmente, en la Figura 3.27 se muestran los Score Plots que relacionan los scores obtenidos para las variables X (denotados por T), frente a los scores obtenidos para las variables Y (denotados por U) para las dos componentes principales. Observando ambos gráficos, se concluye la existencia de una relación lineal entre los scores, lo que deriva en la validez de los resultados obtenidos del PLS.

3.3.3. Interpretación de los resultados

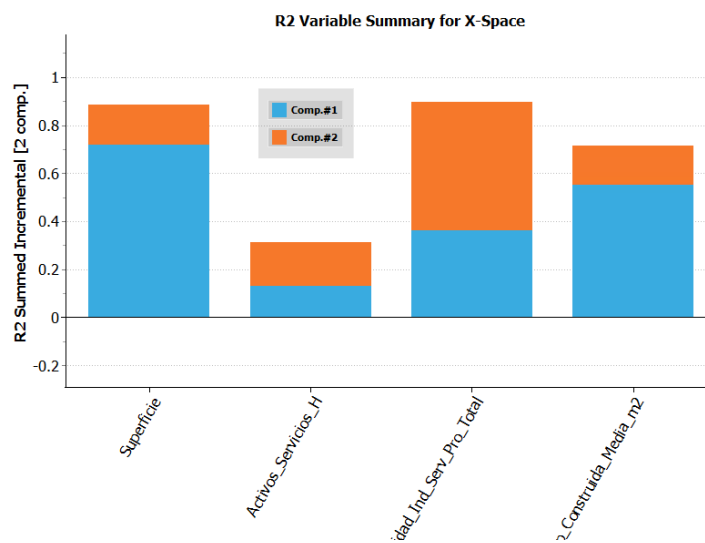
Una vez se ha validado el modelo PLS definitivo, habiendo retenido dos componentes principales que explican el 82% de la variabilidad presente en los datos, examinando la Figura 3.28 se puede determinar que los cinco indicadores de huella de carbono están explicados por la primera componente principal, tal y como se mencionó anteriormente. La segunda componente principal recoge un pequeño porcentaje que la primera componente no es capaz de explicar por sí sola.

Figura 3.28.- R² de cada variable Y del modelo PLS para cada componente principal



Este porcentaje de variabilidad explicada por la segunda componente, tal y como se visualiza en la Figura 3.29, se recoge fundamentalmente en la variable número de habitantes dedicados a actividades industriales y de servicios profesionales y número de hombres que trabajan en el sector servicios. Por tanto, si no se hubiese retenido la segunda componente, estas variables no estarían recogidas en el modelo y faltaría por explicar parte de la información contenida en los datos. Además, merece mencionar que esta última variable es la que tiene una menor importancia en el modelo, ya que su R^2 es considerablemente más reducido en comparación con el resto.

Figura 3.29.- R^2 de cada variable X del modelo PLS para cada componente principal



Finalmente, con el fin de proceder a una completa interpretación de los resultados obtenidos del PLS, en la Figura 3.30 y en la Figura 3.31 se exhibe el Weighting Plot y el Score Plot, respectivamente, del modelo para la componente 1 frente a la 2.

De la información mostrada en ambos gráficos se deriva que los barrios que presentan una mayor huella de carbono total son aquellos con un mayor número de habitantes dedicados a actividades industriales y de servicios profesionales, una mayor extensión y una menor superficie media construida. La variable que determina el número de hombres que trabajan en el sector servicios es la que menor importancia tiene en el modelo, tal y como se adelantaba en la Figura 3.29.

Figura 3.30.- Weighting Plot (W^* , $c_1 - W^*$, c_2) del modelo PLS

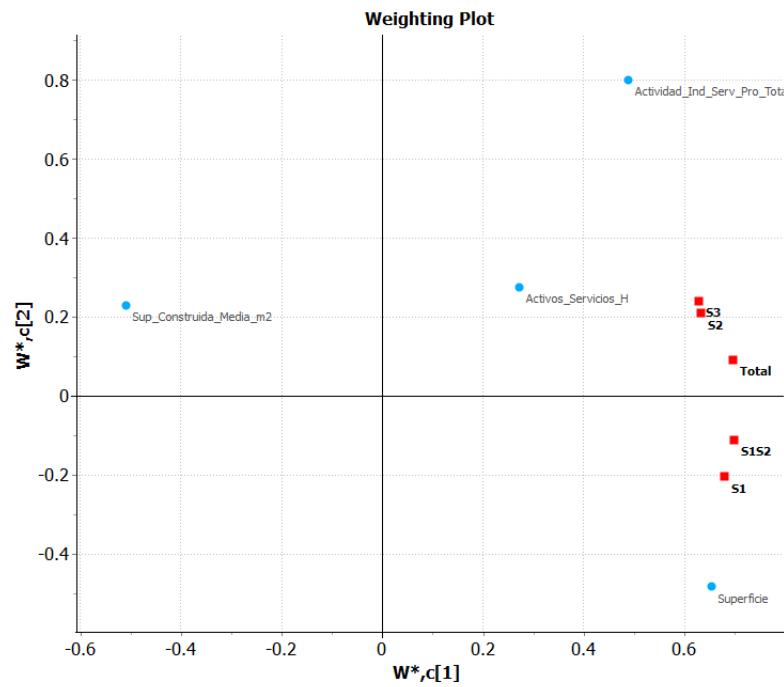
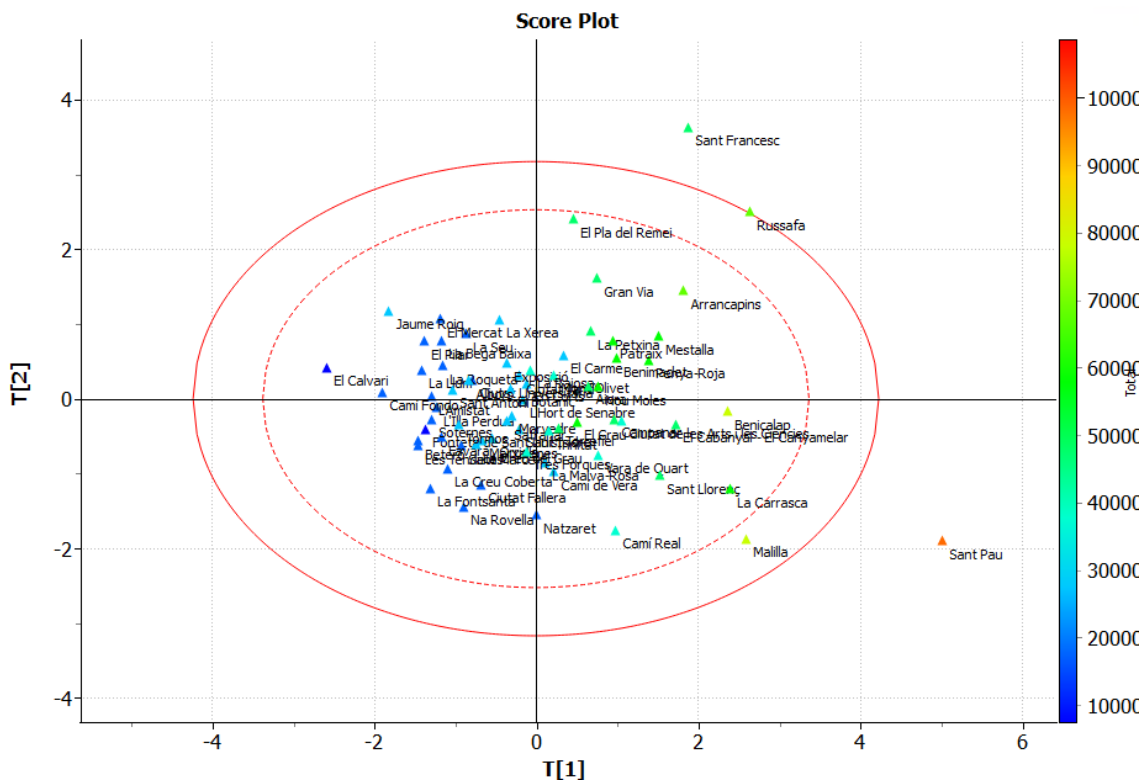


Figura 3.31.- Score Plot ($T_1 - T_2$) del modelo PLS coloreado por la huella de carbono total

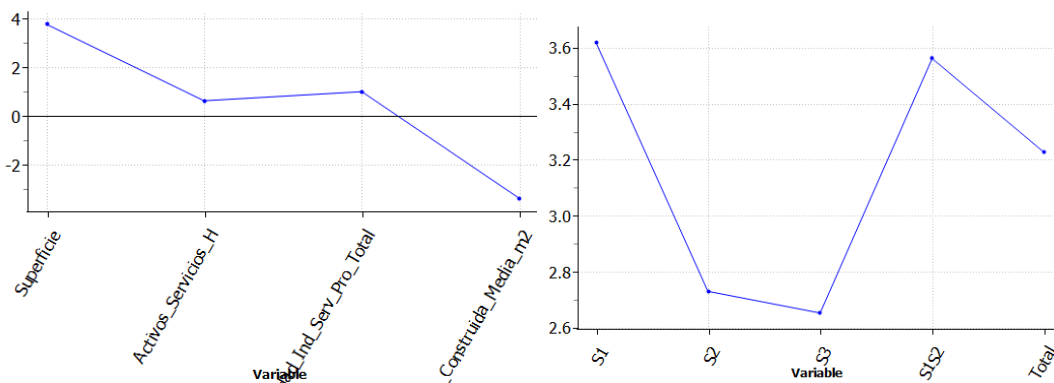


Tras haber coloreado el Score Plot en función de los valores que toma el indicador “Total”, destaca el barrio de Sant Pau como aquel con un valor más elevado (coloreado en naranja). Empleando el Model Explorer del programa Aspen ProMV, en la **Figura 3.32** se observa que la zona por donde se sitúa este barrio (valores elevados y positivos para la

primera componente), se caracteriza por presentar valores más elevados para S1, S1S2 y el Total, y un valor más reducido para S2 y S3.

Asimismo, la Figura 3.32 también refleja la relación negativa que existe entre los indicadores agrupados de huella de carbono y la superficie construida media, y se destaca la variable superficie como aquella que mayor influencia tiene en el análisis.

Figura 3.32.- Gráficos para las variables X (izquierda) e Y (derecha) en la zona con valores elevados en la primera componente



Con toda la información mostrada anteriormente, se puede concluir que las variables socioeconómicas no tienen una relación estadísticamente significativa con los indicadores que miden la huella de carbono de cada barrio. En otras palabras, estas variables no explican por qué algunos barrios tienen una huella de carbono más elevada que otros.

No obstante, entre estos indicadores socioeconómicos se pueden destacar cuatro que sí ejercen influencia y permiten explicar hasta un 82% de la variabilidad contenida en los datos con una bondad de predicción del 79,59%. Estas variables son: la extensión total de cada barrio, el número de ciudadanos dedicados a actividades industriales y de servicios profesionales, el número de varones que trabajan en el sector servicios y la superficie media construida en cada barrio. Tres de estos cuatro indicadores muestran una relación positiva con la huella de carbono total, siendo la última variable mencionada la única que presenta una relación inversa. Esto significa que a medida que aumenta la superficie construida en un barrio, menor es su huella de carbono.

CONCLUSIONES

Uno de los desafíos más importantes en la actualidad es el desarrollo y la implementación de medidas efectivas para combatir el cambio climático. En este contexto, la sostenibilidad urbana ha emergido como una de las prioridades globales.

La realización del presente TFM ha permitido analizar la huella de carbono de los diferentes barrios de la ciudad de Valencia en base a diversos indicadores individuales y agrupados. Gracias al empleo de técnicas de análisis estadístico espacial, se han podido estudiar las relaciones entre los barrios de la ciudad concluyendo que, desagregando los indicadores a nivel individual, existen regiones específicas cuyos resultados se ven influenciados por el entorno en el que se encuentran.

En concreto, cabe destacar que los barrios situados más al sur de la ciudad han presentado una huella de carbono elevada para los indicadores relacionados con los procesos industriales y de construcción, cuyos impactos negativos se relacionan con el entorno en el que se ubican. Del mismo modo, los barrios céntricos han registrado resultados desfavorables para las variables relacionadas con los viajes de larga distancia, el consumo energético en los comercios y las pérdidas de energía, los cuales también se encuentran relacionados por el desempeño de sus vecinos. No obstante, estos barrios han destacado de manera positiva en lo que respecta al consumo de combustibles fósiles y electricidad en transportes de corta distancia.

Por otro lado, los barrios del noroeste de Valencia se han visto resaltados por sus bajas emisiones en indicadores relacionados con la construcción y el consumo de energía en la industria y el comercio, cuyos buenos resultados están vinculados a su entorno. Finalmente, las zonas oeste y suroeste de la ciudad también han mostrado resultados positivos relacionados con el entorno en el que se encuentran, concretamente para los indicadores relacionados con el consumo de energía en los comercios y las pérdidas de energía. Es importante destacar que la absorción de CO₂ por las áreas verdes de cada barrio no ha presentado resultados estadísticamente significativos, lo que significa que no existe una relación espacial entre barrios.

Asimismo, los análisis estadísticos espaciales desarrollados para los indicadores agrupados no han arrojado resultados significativos, lo que refleja que no existe una relación espacial para la huella de carbono total entre los barrios de la ciudad de Valencia.

En otras palabras, los resultados globales que miden la huella de carbono para cada barrio no están relacionados por su cercanía. En consecuencia, para poder observar agrupaciones de barrios estadísticamente significativos, se debe recurrir a la desagregación de la huella de carbono en indicadores individuales.

Utilizando técnicas de análisis multivariante, se han podido identificar clústeres de barrios dispersos por la ciudad con características similares, destacando el método de Ward como aquel que arroja mejores resultados. Además, se ha determinado que la imposición de restricciones espaciales para conectar barrios cercanos con una huella de carbono similar impide la formación de estos grupos, resultados que coinciden con los hallazgos previamente descritos.

Por último, se han estudiado las relaciones entre distintas variables socioeconómicas que miden las características de cada barrio y los indicadores agrupados de huella de carbono. Mediante el empleo del modelo PLS basado en estructuras latentes, se ha llegado a la conclusión de que no existe una relación estadísticamente significativa entre las variables socioeconómicas y los indicadores agrupados de huella de carbono. Esto sugiere que otros factores, posiblemente estructurales o de comportamiento, podrían estar jugando un papel más determinante en las emisiones de CO₂ a nivel de barrio.

Los hallazgos de este estudio proporcionan una herramienta valiosa para los responsables de la toma de decisiones en materia de política ambiental. La identificación de relaciones entre barrios con mayores emisiones y las variables que más influyen en estas puede guiar la implementación de medidas más efectivas y focalizadas a nivel local, promoviendo un desarrollo urbano más sostenible.

Si bien de manera generalizada se ha comprobado que los malos resultados de algunos barrios respecto a su huella de carbono no están relacionados con los de sus vecinos, es decir, no existe una relación espacial significativa, este proyecto ha demostrado la efectividad de las técnicas de análisis estadístico espacial para identificar relaciones entre regiones geográficamente cercanas.

En futuras investigaciones, se plantea la posibilidad de emplear otras técnicas de análisis espacial, como la regresión espacial, para desarrollar modelos predictivos de la huella de carbono en un barrio. Este tipo de modelos incorporarían relaciones espaciales y permitirían comparar su desempeño con los modelos de regresión tradicionales o aquellos basados en estructuras latentes. Asimismo, siguiendo esta última metodología,

sugiere la incorporación de nuevas variables, como son las de tipo conductual y urbanístico, que puedan explicar las diferencias en la huella de carbono entre los barrios de la ciudad.

BIBLIOGRAFÍA

- Afifi, A., May, S., Donatello, R. A., & Clark, V. A. (2019). *Practical Multivariate Analysis*. Chapman and Hall/CRC. <https://doi.org/10.1201/9781315203737>
- Ajuntament de València. (2018). *Plan Especial de Directrices de Calidad Urbana*. https://www.valencia.es/revisionplan/sites/default/files/docs/ped_web.pdf
- Akhter, J., & Afroz, R. (2024). Influence of Climate Variability and Land Cover Dynamics on the Spatio-temporal NDVI Patterns in Western Hydrological Regions of Bangladesh. *Heliyon*. <https://doi.org/10.1016/j.heliyon.2024.e32625>
- Anselin, L. (1980). *Estimation methods for spatial autoregressive structures: a study in spatial econometrics*. Ithaca (N.Y.): Cornell university. Program in urban and regional studies. <http://lib.ugent.be/catalog/rug01:000799976>
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Kluwer Academic Publishers,.
- 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>
- Anselin, L., Florax, R., & Rey, S. J. (2013). *Advances in spatial econometrics: methodology, tools and applications*. Springer Science & Business Media.
- ArcGis Desktop. (n.d.). *Extensiones de archivo de shapefile*.
- AspenTech. (2017). *Aspen ProMV (10.0)*. Product & Process Quality Analytics.
- Ayuntamiento de Valencia. (n.d.). *Oficina de Estadística*.
- Brander, M., & Davis, G. (2012). Greenhouse gases, CO₂, CO₂e, and carbon: What do all these terms mean. *Econometrica, White Papers*.
- Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1–27. <https://doi.org/10.1080/03610927408827101>
- Cliff, A. D., & Ord, J. K. (1973). Spatial autocorrelation. (*No Title*).
- Cliff, A. D., & Ord, J. K. (1981). Spatial processes: models & applications. (*No Title*).
- Duque, J. C., Ramos, R., & Suriñach, J. (2007). Supervised Regionalization Methods: A Survey. *International Regional Science Review*, 30(3), 195–220. <https://doi.org/10.1177/0160017607301605>
- Ermagun, A., & Levinson, D. (2018). Spatiotemporal traffic forecasting: review and proposed directions. *Transport Reviews*, 38(6), 786–814. <https://doi.org/10.1080/01441647.2018.1442887>
- ESRI. (1998). ESRI Shapefile Technical Description. In *ESRI White Paper*.
- European Commission. (2024, May 16). *European Green Capital Award*. Environment.

- European Commission, Directorate-General for Research and Innovation, Gronkiewicz-Waltz, H., Larsson, A., Boni, A., Krogh Andersen, K., Ferrao, P., Forest, E., Jordan, R., Lenz, B., Lumbreras, J., Nicolaidis, C., Reiter, J., Russ, M., Sulling, A., Termont, D., & Vassilakou, M. (2020). *100 climate-neutral cities by 2030 - by and for the citizens – Report of the mission board for climate-neutral and smart cities*. Publications Office. <https://doi.org/doi/10.2777/46063>
- Friedman, J. (2001). *The elements of statistical learning* (2nd ed., Vol. 1). Springer Series in Statistics.
- Geary, R. C. (1954). The contiguity ratio and statistical mapping. *The Incorporated Statistician*, 5(3), 115–146.
- Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: a tutorial. *Analytica Chimica Acta*, 185, 1–17.
- Getis, A. (2007). Reflections on spatial autocorrelation. *Regional Science and Urban Economics*, 37(4), 491–496. <https://doi.org/10.1016/j.regsciurbeco.2007.04.005>
- Getis, A., & Ord, J. K. (1992). The Analysis of Spatial Association by Use of Distance Statistics. *Geographical Analysis*, 24(3), 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Ghodousi, M., Sadeghi-Niaraki, A., Rabiee, F., & Choi, S.-M. (2020). Spatial-Temporal Analysis of Point Distribution Pattern of Schools Using Spatial Autocorrelation Indices in Bojnourd City. *Sustainability*, 12(18). <https://doi.org/10.3390/su12187755>
- Gillies, S. (2024). *The Shapely User Manual* (2.0.4).
- Gómez Rubio, V. (2022). Una introducción a la estadística espacial. *Boletín Oficial de Estadística e Investigación Operativa*, 38(1).
- Greenhouse Gas Protocol. (2022). *Protocolo Global para Inventarios de Emisión de Gases de Efecto Invernadero a Escala Comunitaria*. https://ghgprotocol.org/sites/default/files/2022-12/GHGP_GPC%20%28Spanish%29.pdf
- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, 585(7825), 357–362. <https://doi.org/10.1038/s41586-020-2649-2>
- Horn, D. L., Hampton, C. R., & Vandenberg, A. J. (1993). Practical application of district compactness. *Political Geography*, 12(2), 103–120. [https://doi.org/10.1016/0962-6298\(93\)90031-2](https://doi.org/10.1016/0962-6298(93)90031-2)
- Hunter, J. D. (2007). Matplotlib: A 2D graphics environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- INE. (2023). *Cifras oficiales de población de los municipios españoles: Revisión del Padrón Municipal*. Instituto Nacional de Estadística.

- Jordahl, K., den Bossche, J. Van, Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., ... Leblanc, F. (2020). *geopandas/geopandas: v0.8.1*. Zenodo. <https://doi.org/10.5281/zenodo.3946761>
- Kaufman, L., & Rousseeuw, P. J. (1990). *Finding Groups in Data: An Introduction To Cluster Analysis*. Wiley. <https://doi.org/10.1002/9780470316801>
- Keshtkar, M., Heidari, H., Moazzeni, N., & Azadi, H. (2022). Analysis of changes in air pollution quality and impact of COVID-19 on environmental health in Iran: application of interpolation models and spatial autocorrelation. *Environmental Science and Pollution Research*, 29(25), 38505–38526. <https://doi.org/10.1007/s11356-021-17955-9>
- Kumar, K. (2021). Partial Least Square (PLS) Analysis. *Resonance*, 26(3), 429–442. <https://doi.org/10.1007/s12045-021-1140-1>
- Lamb, D. S., Downs, J., & Reader, S. (2020). Space-Time Hierarchical Clustering for Identifying Clusters in Spatiotemporal Point Data. *ISPRS International Journal of Geo-Information*, 9(2). <https://doi.org/10.3390/ijgi9020085>
- Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129–137. <https://doi.org/10.1109/TIT.1982.1056489>
- Lombardi, M., Laiola, E., Tricase, C., & Rana, R. (2017). Assessing the urban carbon footprint: An overview. *Environmental Impact Assessment Review*, 66, 43–52. <https://doi.org/10.1016/j.eiar.2017.06.005>
- Lombardi, M., Laiola, E., Tricase, C., & Rana, R. (2018). Toward urban environmental sustainability: The carbon footprint of Foggia’s municipality. *Journal of Cleaner Production*, 186, 534–543. <https://doi.org/10.1016/j.jclepro.2018.03.158>
- Mardia, K. V, Kent, J. T., & Taylor, C. C. (2024). *Multivariate analysis* (Vol. 88). John Wiley & Sons.
- Mazzucato, M. (2018). *Mission-oriented research & innovation in the European Union – A problem-solving approach to fuel innovation-led growth*. European Commission – Directorate-General for Research and Innovation. <https://doi.org/doi/10.2777/360325>
- Melecky, L. (2015). Spatial Autocorrelation Method for Local Analysis of The EU. *Procedia Economics and Finance*, 23, 1102–1109. [https://doi.org/10.1016/S2212-5671\(15\)00484-0](https://doi.org/10.1016/S2212-5671(15)00484-0)
- Miller, S. A., Monteiro, P. J. M., Ostertag, C. P., & Horvath, A. (2016). Comparison indices for design and proportioning of concrete mixtures taking environmental impacts into account. *Cement and Concrete Composites*, 68, 131–143. <https://doi.org/10.1016/j.cemconcomp.2016.02.002>
- Missions València 2030. (2024). *Propuestas de misiones*. <https://www.missionsvalencia.eu/missions/?lang=es>

- Moran, P. A. P. (1948). The Interpretation of Statistical Maps. *Journal of the Royal Statistical Society Series B-Methodological*, 10, 243–251. <https://api.semanticscholar.org/CorpusID:134815800>
- Moreno, C., Allam, Z., Chabaud, D., Gall, C., & Pratlong, F. (2021). Introducing the “15-Minute City”: Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities*, 4(1), 93–111. <https://doi.org/10.3390/smartcities4010006>
- Moreno Serrano, R., & Vayá-Valcarce, E. (2000). *Técnicas econométricas para el tratamiento de datos espaciales: La econometría espacial*. Universitat de Barcelona.
- Mukim, M., & Roberts, M. (2023). *Thriving: Making Cities Green, Resilient, and Inclusive in a Changing Climate*. World Bank.
- Ord, J. K., & Getis, A. (1995). Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis*, 27(4), 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>
- Paelinck, J. (1978). Spatial econometrics. *Economics Letters*, 1(1), 59–63. <https://EconPapers.repec.org/RePEc:eee:ecole:v:1:y:1978:i:1:p:59-63>
- Pearson, K. (1901). On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572. <https://doi.org/10.1080/14786440109462720>
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel V. and Thirion, B., Grisel, O., Blondel, M., Prettenhofer P. and Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.
- Rey, S., Arribas-Bel, D., & Wolf, L. J. (2023). *Geographic Data Science with Python*. Chapman and Hall/CRC. <https://doi.org/10.1201/9780429292507>
- Rey, S. J., & Anselin, L. (2007). PySAL: A Python Library of Spatial Analytical Methods. *The Review of Regional Studies*, 37(1), 5–27.
- Roncero Tarazona, M., Crespo Montañés, C., Rodolfo Martínez, L., Ribó Pérez, D., Chong Gil, P., Gómez Navarro, T., & Montagud Montalvá, C. (2024). *Enabling Carbon Footprint accounting at neighbourhood level. Case of application to a Mediterranean compact city*. (in press)
- Sammut, C., & Webb, G. I. (2010). *Encyclopedia of Machine Learning* (C. Sammut & G. I. Webb, Eds.). Springer US. <https://doi.org/10.1007/978-0-387-30164-8>
- 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>
- Sorli Guerola, J. (2023, January 27). *El Racó de l'Olla*. Birdingplaces.
- Standen, R. van. (2014). *Climate Change: Implications for Cities*.

- Sun, M., Liang, B., Meng, X., Zhang, Y., Wang, Z., & Wang, J. (2024). Study on the Evolution of Spatial and Temporal Patterns of Carbon Emissions and Influencing Factors in China. *Land*, 13(6), 828. <https://doi.org/10.3390/land13060828>
- The pandas development team. (2024). *pandas-dev/pandas: Pandas (2.2.2)*. Zenodo. <https://doi.org/https://doi.org/10.5281/zenodo.10957263>
- Tobler, W. R. (1970). A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, 46, 234. <https://doi.org/10.2307/143141>
- United Nations. (2015). *Resolution adopted by the General Assembly on 25 September 2015*.
- United Nations Climate Change. (n.d.). *Protocolo de Kioto*.
- Van Rossum, G. (2024). *Python (3.12.4)*.
- Varghese, B. M., Unnikrishnan, A., & Jacob, K. (2013). Spatial clustering algorithms-an overview. *Asian Journal of Computer Science and Information Technology*, 3(1), 1–8.
- Vinh, N. X., Epps, J., & Bailey, J. (2009). Information theoretic measures for clusterings comparison. *Proceedings of the 26th Annual International Conference on Machine Learning*, 1073–1080. <https://doi.org/10.1145/1553374.1553511>
- Vinzi, V. E., Chin, W. W., Henseler, J., & Wang, H. (2010). *Handbook of partial least squares* (Vol. 201, Issue 0). Springer.
- Waskom, M. (2021). seaborn: statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>
- WBCSD. (2024). *World Business Council for Sustainable Development*. <https://www.wbcsd.org/>
- Webber, R., & Burrows, R. (2018). *The Predictive Postcode The Geodemographic Classification of British Society*. Sage.
- Wold, H. (1979). Model construction and evaluation when theoretical knowledge is scarce: An example of the use of partial least squares cahiers du departement deconometrie. *Geneva, Switzerland: Faculté Des Sciences Économiques et Sociales, Université de Genève*.
- Wold, H. (1982). Soft modelling: the basic design and some extensions. *Systems under Indirect Observation, Part II*, 36–37.
- Wold, S., Esbensen, K., & Geladi, P. (1987). Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, 2, 37–52. [https://doi.org/10.1016/0169-7439\(87\)80084-9](https://doi.org/10.1016/0169-7439(87)80084-9)
- Wold, S., Ruhe, A., Wold, H., & Dunn WJ, I. (1984). The collinearity problem in linear regression. The partial least squares (PLS) approach to generalized inverses. *SIAM Journal on Scientific and Statistical Computing*, 5(3), 735–743.

- Wold, S., Sjöström, M., & Eriksson, L. (2001). PLS-regression: a basic tool of chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 58(2), 109–130. [https://doi.org/10.1016/S0169-7439\(01\)00155-1](https://doi.org/10.1016/S0169-7439(01)00155-1)
- World Resources Institute. (2024). *World Resources Institute | Making Big Ideas happen*. <https://www.wri.org/>
- Yan, Z., Sun, Z., Shi, R., & Zhao, M. (2023). Smart city and green development: Empirical evidence from the perspective of green technological innovation. *Technological Forecasting and Social Change*, 191, 122507. <https://doi.org/10.1016/j.techfore.2023.122507>
- Yang, F., Yu, J., Zhang, C., Li, L., Lei, Y., Wu, S., Wang, Y., & Zhang, X. (2024). Spatio-temporal differentiation characteristics and the influencing factors of PM2.5 emissions from coal consumption in Central Plains Urban Agglomeration. *Science of The Total Environment*, 173778. <https://doi.org/10.1016/j.scitotenv.2024.173778>

ANEXO I

Grado de relación del Trabajo Fin de Máster con los Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030.

Objetivos de desarrollo sostenible	Alto	Medio	Bajo	No procede
ODS 1. Fin de la pobreza.				X
ODS 2. Hambre cero.				X
ODS 3. Salud y bienestar.		X		
ODS 4. Educación de calidad.				X
ODS 5. Igualdad de género.				X
ODS 6. Agua limpia y saneamiento.			X	
ODS 7. Energía asequible y no contaminante.			X	
ODS 8. Trabajo decente y crecimiento económico.				X
ODS 9. Industria, innovación e infraestructuras.				X
ODS 10. Reducción de las desigualdades.				X
ODS 11. Ciudades y comunidades sostenibles.	X			
ODS 12. Producción y consumo responsables.			X	
ODS 13. Acción por el clima.	X			
ODS 14. Vida submarina.				X
ODS 15. Vida de ecosistemas terrestres.				X
ODS 16. Paz, justicia e instituciones sólidas.				X
ODS 17. Alianzas para lograr objetivos.				X

El presente TFM tiene un grado de relación alto con dos Objetivos de Desarrollo Sostenible, tal y como se relató al inicio del estudio. En línea con la *Misión Climática València 2030*, cuyo propósito es conseguir que la ciudad sea climáticamente neutra para 2030, el ODS número 11: “Ciudades y comunidades sostenibles” se encuentra en el punto de mira. Con la realización de este proyecto se ofrece una perspectiva de la situación de la ciudad de Valencia en términos de emisiones de CO₂ a nivel de barrio. Por medio de técnicas estadísticas de análisis espacial y análisis multivariante, se estudian las relaciones entre la huella de carbono de cada barrio con el fin de detectar regiones cuyos resultados se vean influenciados por las emisiones de otras áreas. En consecuencia, este proyecto también se encuentra estrechamente relacionado con el ODS número 13: “Acción por el clima”.

En conjunto, todos los análisis presentados proporcionan una herramienta de ayuda para las entidades responsables, permitiéndoles orientar sus estrategias y decisiones basándose en los resultados obtenidos.

ANEXO II

VARIABLES SOCIOECONÓMICAS EMPLEADAS EN EL ESTUDIO.

Poblacion	Edu_Menor_18	TB_Emig_Interurbana	Activos_Cooperativa_Total
Dens_Pob	Edu_Menor_18_H	TB_Fecundidad	Activos_Cooperativa_H
Pob_Hombres	Edu_Menor_18_M	Solteros	Activos_Cooperativa_M
Pob_Mujeres	Edu_Analfabeto	Solteros_H	Activos_Paro_sin_Trabajo_antes_Total
Edad_0_4	Edu_Analfabeto_H	Solteros_M	Activos_Paro_sin_Trabajo_antes_H
Edad_0_4_H	Edu_Analfabeto_M	Casados	Activos_Paro_sin_Trabajo_antes_M
Edad_0_4_M	Edu_Inf_Grad_Escolar	Casados_H	Actividad_Ind_Serv_Pro_Total
Edad_5_9	Edu_Inf_Grad_Escolar_H	Casados_M	Actividad_Ind_Total
Edad_5_9_H	Edu_Inf_Grad_Escolar_M	Viudos	Actividad_Ind_Energia_Agua
Edad_5_9_M	Inmig_America_Central	Viudos_H	Actividad_Ind_Ext_Minerales
Edad_10_14	Inmig_America_Sur	Viudos_M	Actividad_Ind_Metales_Mec_Precision
Edad_10_14_H	Inmig_Asia	Separados	Actividad_Ind_Otros
Edad_10_14_M	Inmig_Oceania_Otros	Separados_H	Actividad_Serv_Total
Edad_15_19	Edu_Menor_18	Separados_M	Actividad_Serv_Comercio_Hosteleria
Edad_15_19_H	Edu_Menor_18_H	Divorciados	Actividad_Serv_Transporte_Teleco
Edad_15_19_M	Edu_Menor_18_M	Divorciados_H	Actividad_Serv_Fin_Seg_Alq
Edad_20_24	Edu_Analfabeto	Divorciados_M	Actividad_Serv_Otros
Edad_20_24_H	Edu_Analfabeto_H	Mujeres_sin_Hijos	Actividad_Prof_Total
Edad_20_24_M	Edu_Analfabeto_M	Mujeres_con_Hijos	Actividad_Prof_Agricultura_Ganaderia
Edad_25_29	Edu_Inf_Grad_Escolar	Mujeres_1_Hijo	Actividad_Prof_Energia_Mineria
Edad_25_29_H	Edu_Inf_Grad_Escolar_H	Mujeres_2_Hijo	Actividad_Prof_Aero_Comm_Mec
Edad_25_29_M	Edu_Inf_Grad_Escolar_M	Mujeres_3_Hijo	Actividad_Prof_Ind_Otras
Edad_30_34	Edu_Grad_Escolar	Mujeres_4_mas_Hijo	Actividad_Prof_Construccion
Edad_30_34_H	Edu_Grad_Escolar_H	Mayores_16_Total	Actividad_Prof_Comercio_Hosteleria
Edad_30_34_M	Edu_Grad_Escolar_M	Mayores_16_H	Actividad_Prof_Comm_Transporte
Edad_35_39	Edu_Bach_mas	Mayores_16_M	Actividad_Prof_fina_jud_seg_alq
Edad_35_39_H	Edu_Bach_mas_H	Activos_Total	Actividad_Prof_Otros
Edad_35_39_M	Edu_Bach_mas_M	Activos_H	Vehiculos_Total
Edad_40_44	Tam_1_Hoja_Fam	Activos_M	Vehiculos_Turismo
Edad_40_44_H	Tam_2_Hoja_Fam	Ocupados_T.Completo_Total	Vehiculos_Autobus

Edad_40_44_M	Tam_3_Hoja_Fam	Ocupados_T.Completo_H	Vehiculos_Camion
Edad_45_49	Tam_4_Hoja_Fam	Ocupados_T.Completo_M	Vehiculos_Tractor
Edad_45_49_H	Tam_5_Hoja_Fam	Ocupados_T.Parcial_Total	Vehiculos_Remolque
Edad_45_49_M	Tam_6_Hoja_Fam	Ocupados_T.Parcial_H	Vehiculos_Motocicleta
Edad_50_54	Tam_7_mas_Hoja_Fam	Ocupados_T.Parcial_M	Vehiculos_Ciclomotor
Edad_50_54_H	Menores_18_0_Hoja_Fam	Parados_Trabajo_antes_Total	Turismos_0_8_cv
Edad_50_54_M	Menores_18_1_Hoja_Fam	Parados_Trabajo_antes_H	Turismos_8_12_cv
Edad_55_59	Menores_18_2_Hoja_Fam	Parados_Trabajo_antes_M	Turismos_12_16_cv
Edad_55_59_H	Menores_18_3_Hoja_Fam	Parados_Primer_Trabajo_T	Turismos_16_20_cv
Edad_55_59_M	Menores_18_4_Hoja_Fam	Parados_Primer_Trabajo_H	Turismos_20_mas_cv
Edad_60_64	Menores_18_5_mas_Hoja_Fam	Parados_Primer_Trabajo_M	Turismos_Particulares
Edad_60_64_H	Edad_Alguien_16_24_Hoja_Fam	Inactivos_Total	Turismos_Empresa_Org
Edad_60_64_M	Edad_Alguien_25_64_Hoja_Fam	Inactivos_H	Viviendas_Total
Edad_65_69	Edad_Alguien_mas_64_Hoja_Fam	Inactivos_M	Viviendas_1700_1800
Edad_65_69_H	Edad_Solo_mas_64_Hoja_Fam	Activos_AGP_Total	Viviendas_1801_1900
Edad_65_69_M	Edad_Alguien_mas_79_Hoja_Fam	Activos_AGP_Hom	Viviendas_1901_1920
Edad_70_74	Edad_Solo_mas_79_Hoja_Fam	Activos_AGP_Muj	Viviendas_1921_1940
Edad_70_74_H	Edad_Media	Activos_Industria_Total	Viviendas_1941_1960
Edad_70_74_M	Indice_Envejecim	Activos_Industria_H	Viviendas_1961_1980
Edad_75_79	Indice_Sobreevejecim	Activos_Industria_M	Viviendas_1981_2000
Edad_75_79_H	Indice_Demo_Dependencia	Activos_Construccion_Total	Viviendas_2001_2010
Edad_75_79_M	Indice_Estructura_Pob_Activa	Activos_Construccion_H	Viviendas_2011_act
Edad_80_84	Indice_Reemplazamiento_Pob_Activa	Activos_Construccion_M	Viviendas_menos_60_m2
Edad_80_84_H	Razon_Progresividad_Demografica	Activos_Servicios_Total	Viviendas_61_80_m2
Edad_80_84_M	Alta_Inmigracion	Activos_Servicios_H	Viviendas_81_100_m2
Edad_85_mas	Alta_Nacimiento	Activos_Servicios_M	Viviendas_101_120_m2
Edad_85_mas_H	Alta_Cambio_Domicilio	Activos_Empresario_con_Personal_Total	Viviendas_121_150_m2
Edad_85_mas_M	Alta_Otros	Activos_Empresario_con_Personal_H	Viviendas_151_200_m2
Inmig_Est_Total	Alta_Total	Activos_Empresario_con_Personal_M	Viviendas_201_mas_m2
Inmig_Est_H	Baja_Emigracion	Activos_Empresario_sin_Personal_Total	Sup_Construida_Media_m2
Inmig_Est_M	Baja_Defuncion	Activos_Empresario_sin_Personal_H	Viviendas_Valor_Cat_0_12
Inmig_Ext_Total	Baja_Cambio_Domicilio	Activos_Empresario_sin_Personal_M	Vivivedas_Valor_Cat_12_18

Inmig_Ext_H	Baja_Otros	Activos_Empresario_sin_Personal_Total	Viviendas_Valor_Cat_18_24
Inmig_Ext_M	Baja_Total	Activos_Empresario_sin_Personal_H	Viviendas_Valor_Cat_24_30
Inmig_Europa	Crecimiento_Vegetativo	Activos_Empresario_sin_Personal_M	Viviendas_Valor_Cat_30_36
Inmig_Europa_U_E	Saldo_Intraurbano	Activos_Empleados_Fijos_Total	Viviendas_Valor_Cat_36_48
Inmig_Europa_N_o_UE	Saldo_Interurbano	Activos_Empleados_Fijos_H	Viviendas_Valor_Cat_48_60
Inmig_Africa	Otros_Cambios	Activos_Empleados_Fijos_M	Viviendas_Valor_Cat_60_72
Inmig_America	Saldo_Neto	Activos_Empleados_Eventuales_Total	Viviendas_Valor_Cat_72_mas
Inmig_America_Norte	TB_Natalidad	Activos_Empleados_Eventuales_H	Viviendas_Valor_Cat_NA
Inmig_America_Central	TB_Inmig_Intraurbana	Activos_Empleados_Eventuales_M	Valor_Suelo
Inmig_America_Sur	TB_Inmig_Interurbana	Activos_Ayuda_Familiar_T	Valor_Construccion
Inmig_Asia	TB_Mortalidad	Activos_Ayuda_Familiar_H	Valor_Total
Inmig_Oceania_Otros	TB_Emig_Intraurbana	Activos_Ayuda_Familiar_M	Valor_m2
	Sup_Parking_Viv	Sup_Parking_Tur	Sup_Parking_Total_m2