



IDENTIFICACIÓN DE PUNTOS DE INTERÉS TURÍSTICO A TRAVÉS DE DATOS DE REDES SOCIALES

Valls, Francesc ^{1*}; Roca, Josep ²

Remisión inicial: 2019-07-11; **Remisión definitiva:** 2019-10-22; **Publicación:** 2019-12-21

Citación: Valls, F. y Roca, J. (2019). Identificación de puntos de interés turístico a través de datos de redes sociales. En *XIII CTV 2019 Proceedings: XIII International Conference on Virtual City and Territory: "Challenges and paradigms of the contemporary city"*. UPC, Barcelona, October 2-4, 2019. Barcelona: CPSV, 2019, p. 8755. E-ISSN 2604-6512. DOI <http://dx.doi.org/10.5821/ctv.8755>

Resumen

En muchas ciudades europeas, el turismo de masas se está convirtiendo en una importante industria, pero los efectos de la creciente actividad de los visitantes están empezando a ser percibidas de manera negativa por sus habitantes. Las autoridades locales deben implementar políticas que modulen estas actividades para que puedan coexistir y ser compatibles con la vida diaria de la población residente. Sin embargo, estas políticas deben tener en consideración que la actividad turística tiende a concentrarse en áreas muy específicas, las cuales deben ser identificadas claramente antes de imponer formas específicas de tasación, o requerir la obtención de una licencia obligatoria para ciertas actividades. Para delimitar de manera clara estos ámbitos, esta investigación propone utilizar los datos de las redes sociales para identificar puntos calientes de actividad turística en Barcelona, como una fuente emergente de información en la toma de decisiones, utilizando como estudio de caso más de 75.000 fotografías geolocalizadas obtenidas del servicio Panoramio a través de su interfaz de programación de aplicaciones (API).

Esta aproximación al análisis urbano basada en datos debe solventar algunos aspectos de lo que algunos autores describen como "las 4 V del Big Data": volumen, variedad, velocidad, y veracidad. En particular, esta investigación es especialmente sensible a dos aspectos (volumen y veracidad) que debían ser resueltos de manera efectiva. En el caso del volumen, el gran número de localizaciones dificultaba la aplicación de las técnicas de análisis espacial convencional, mientras que, en el caso de la veracidad, la naturaleza informal de los datos reducía la confianza en la precisión de su ubicación. La metodología siguió una aproximación basada en la estadística espacial, centrada en la distribución espacial de las localizaciones analizada a través en la estimación de la densidad por núcleo (KDE), con el ancho de banda determinado con el objetivo de identificar tendencias generales a la escala de estudio y reducir el ruido espacial.

Sin embargo, el análisis de la densidad de fotografías de manera aislada ignora la variación espacial de la intensidad de uso dentro de la ciudad, puesto que en áreas de mayor actividad es esperable una mayor concentración de fotografías, *ceteris paribus*. La estimación del nivel de actividad se realizó en base al inventario de locales de la ciudad de Barcelona, considerando que se trataba de un buen indicador de la intensidad de uso de los distintos ámbitos de la ciudad, bajo la premisa que una mayor presencia de locales es indicativa de una mayor actividad peatonal.

Para analizar el atractivo turístico teniendo en cuenta esta heterogeneidad espacial en el uso de la ciudad, se comparó la distribución espacial de fotografías en relación con la de actividad comercial en planta baja (como indicador indirecto de intensidad de uso). Para hacer equiparables ambas distribuciones, se desarrolló una metodología para normalizar los valores obtenidos del KDE, obteniendo un indicador (atractivo relativo) robusto respecto tanto a la resolución de discretización, el número de localizaciones y su distribución espacial. Las diferencias entre ambas superficies normalizadas fueron clasificadas en una escala divergente para identificar y cuantificar los puntos de actividad turística tanto fríos como calientes, permitiendo identificar las áreas de elevada presión turística y también los "desiertos" con muy escasa presencia de visitantes.

El enfoque propuesto propone una vía emergente de investigación en un campo de estudio tradicionalmente escaso en datos, y sugiere que las redes sociales son capaces de convertirse en una valiosa fuente de datos en la investigación urbana. Sin embargo, aunque la cantidad de datos disponible no tiene precedentes, también requiere nuevas técnicas

¹ Dr. Arquitecto, Departamento de Representación Arquitectónica (RA), Universitat Politècnica de Catalunya (UPC), <https://orcid.org/0000-0002-9400-5659>; ² Dr. Arquitecto, Departamento de Tecnología de la Arquitectura (TA), Centro de Política de Suelo y Valoraciones (CPSV), Universitat Politècnica de Catalunya (UPC), <https://orcid.org/0000-0003-3970-6505>. *Correo de contacto: francesc.valls@upc.edu



de análisis, así como el conocimiento específico de los procesos de recopilación, limpieza, análisis y visualización de datos, para obtener resultados relevantes de forma rigurosa.

Abstract

In most European cities, mass tourism is becoming a major industry, but the effects of the increasing visitor pressure are beginning to be negatively perceived by its inhabitants. Local authorities must implement policies that modulate these activities so they can coexist and become compatible with the daily life of the local population. However, these policies must take into consideration that tourist activity tends to concentrate in very specific areas, which must be clearly identified before implementing specific taxation schemes, or requiring the procurement of a mandatory license for certain activities. To clearly designate these areas, this research proposes using social media data to identify hot spots of visitor activity in Barcelona, as an emerging source of information for stakeholders in their decision-making process, using as a case of study more than 75,000 geotagged pictures collected the Panoramio picture sharing community through its Application Programming Interface.

This data-driven approach to urban analysis must address some of the issues of what some authors describe as “the 4 V’s of Big Data”: volume, variety, velocity, and veracity. In particular, this research is especially sensitive to two aspects (volume and veracity) that must be addressed accordingly. In the case of its volume, the large number of locations made conventional spatial data analysis challenging, while in the case of veracity, the informal nature of the source data reduced the confidence on location precision. The methodology followed a principled approach based on spatial statistics, focusing on the spatial distribution of locations analyzed using kernel density estimation (KDE), with the bandwidth determined to identify general trends at the intended scale of analysis and reduce spatial noise.

However, the analysis of the density of photographs in isolation ignores the spatial variation of the intensity of use within the city, since a greater concentration of photographs is expected in areas of elevated activity, *ceteris paribus*. The estimation of the level of activity was based on the inventory of businesses in the city of Barcelona, considering that it was a good indicator of the intensity of use in the different areas of the city, under the premise that a higher number of businesses is indicative increased pedestrian activity.

To analyze tourist attractiveness considering this spatial heterogeneity in the use of the city, the spatial distributions of photographs and street-level commercial activities (as an indirect indicator of intensity of use) were compared. To make both distributions equivalent, a methodology was developed to normalize the values obtained from the KDE, obtaining an indicator (relative attractiveness) robust regarding the resolution of discretization, the number of locations, and their spatial distribution. The differences between both standardized surfaces were classified on a divergent scale to identify and quantify the hot and cold spots of tourist activity, highlighting the areas of outstanding tourist pressure and also the “deserts” almost devoid of visitors.

The proposed approach proposes an emerging avenue of research in a traditionally data-scarce field of study, and suggests that social media is capable of becoming a valuable source of data in urban research. However, while this amount of data available is unprecedented, it also requires new analysis techniques as well as specific domain knowledge in the data collection, cleaning, analysis and visualization processes to successfully provide accurate and meaningful results.

Palabras Clave: redes sociales, turismo, Big Data, análisis espacial urbano

Key words: social media, tourism, Big Data, urban spatial analysis

1. Objetivos

En la ciudad de Barcelona (Mansilla y Milano, 2019), como en muchas otras ciudades históricas europeas, la presión del turismo de masas sobre los centros históricos (García-Hernández, De la Calle-Vaquero, y Yubero, 2017) empieza a ser percibida por sus habitantes de manera negativa. Es por lo tanto crucial implementar políticas de gestión de este turismo para hacerlo compatible con las necesidades de la vida de los habitantes de la ciudad. Sin embargo, estas políticas deben centrarse en las áreas de mayor afluencia, para implementar políticas que modulen este uso a través de la concesión de licencias o la imposición de tasas. Em este sentido es crucial la adecuada identificación y delimitación de las áreas de mayor interés turístico.



Sin embargo, los estudios urbanos han sido tradicionalmente un campo de estudio con importantes limitaciones en la obtención y manejo de datos para su desarrollo. Esta investigación propone una metodología para utilizar los datos de redes sociales como fuente emergente de información (Valls, 2019), con el objetivo de delimitar puntos calientes de actividad turística usando fotografías geolocalizadas.

Esta aproximación basada en datos requiere un nuevo planteamiento en los estudios urbanos, resumidos en los retos que suponen lo que se denomina las tres “V” del “Big Data”: volumen, variedad y velocidad, a los que algunos autores añaden un cuarto concepto al que denominan validez o veracidad (Goodchild, 2016). En este sentido, el objetivo secundario del estudio es desarrollar una metodología robusta capaz de ser replicada, siguiendo los principios de la investigación reproducible (Peng, 2011).

2. Metodología

El presente estudio propone estimar el interés turístico a través de las fotografías geolocalizadas obtenidas de redes sociales, usando como caso de estudio los datos de la red Panoramio. Esta estrategia ha sido utilizada con éxito en distintos estudios recientes (Salas-Olmedo, Moya-Gómez, García-Palomares, y Gutiérrez, 2018), resultando adecuada para el estudio comparativo de múltiples ciudades (García-Palomares, Gutiérrez, y Mínguez, 2015), con un coste económico más reducido y un empleo de tiempo en la recogida de datos más eficiente respecto a otros métodos (Shoval y Ahas, 2016).

Sin embargo, la mayoría de estudios de la literatura analizan la densidad de fotografías de manera aislada, y no tienen en cuenta la distinta distribución espacial de los usos de la ciudad (Anselin y Williams, 2015). De esta manera se ignora que, en áreas de mayor actividad, es esperable una mayor concentración de fotografías, por el simple hecho de la existencia de una mayor presencia de personas. En consecuencia, para estudiar el atractivo turístico teniendo en cuenta esta heterogeneidad en el uso de la ciudad, se compararon las distribuciones espaciales de fotografías en relación con los locales en planta baja.

En la investigación, se partió de dos asunciones principales, respecto a la variable de estudio y a su covariable. La primera asunción supuso que el comportamiento fotográfico de los turistas y los residentes es parecido, partiendo del hecho que las fotografías de Panoramio estaban orientadas a la fotografía de monumentos (*landmarks*), y por lo tanto no se distinguieron perfiles de usuario según su origen; cabe destacar que esta hipótesis no sería válida para otras fuentes de datos orientadas a compartir fotografías de muy distinta índole. La segunda asunción consideró que la intensidad del uso en los distintos ámbitos de la ciudad se puede deducir indirectamente a través de la densidad de locales en planta baja, suponiendo que el mercado inmobiliario sigue una lógica de proveer un mayor número de servicios en los ámbitos donde la actividad es mayor.

3. Datos utilizados

3.1 Datos de redes sociales

Panoramio era una plataforma para compartir fotografías geolocalizadas, fundada por dos emprendedores españoles y lanzada en octubre de 2005. El servicio fue posteriormente



adquirido por Google dos años más tarde y cesó operaciones en noviembre de 2016. A pesar de ello, la web SightsMap² conserva todavía un mapa de calor de las fotografías a escala mundial, mientras que el sitio web MapSights³ tiene como objetivo la preservación estas fotografías. Aunque la popularidad de Panoramio fue siempre menor que otros servicios de compartición de fotografías como Flickr, su enfoque hacia la geolocalización precisa de las fotos la convierte en una fuente de datos valiosa para estudiar el atractivo fotográfico de la ciudad.

A diferencia de la mayoría de redes sociales, Panoramio no requería registro ni autenticación, y tampoco utilizaba encriptación, resultando por lo tanto de más fácil acceso que otros servicios, además de permitir el uso gratuito para uso comercial y no comercial, con un límite de 100.000 peticiones diarias. Panoramio ofrecía dos Web Application Programming Interfaces⁴ (Web APIs): Panoramio Data API⁵, que permitía la descarga de fotos para mostrarlas incrustadas en un mapa dentro de una web externa (por ejemplo, la página de un hotel), y Panoramio Widget API⁶, una biblioteca de JavaScript que permitía incluir un control (widget) en una página web, con un interfaz de usuario y capacidad de búsqueda.

Los datos de Panoramio fueron recogidos utilizando la API de Panoramio (Data API), que permitía especificar distintos parámetros para encontrar las fotografías deseadas, entre las que se encontraban los límites geográficos (*bounding box*) dentro de los cuales se debían buscar las fotografías (Tabla 1), en el sistema de referencia global WGS (World Geodetic System) para la información geoespacial (EPSG:4326), utilizado por el sistema GPS de navegación por satélite.

Tabla 1. **Coordenadas de los límites geográficos en la petición de datos a API de Panoramio**

Extremo	Posición	Grados decimales	DMS
Latitud Mínima	Inferior (Sur)	41,306105095 N	41° 18' 21,98" N
Latitud Máxima	Superior (Norte)	41,478298448 N	41° 28' 41,87" N
Longitud Mínima	Izquierda (Este)	2,044216148 E	2° 2' 39,18" E
Longitud Máxima	Derecho (Oeste)	2,237585016 E	2° 14' 15,31" E

Fuente: Elaboración propia.

Todo el proceso de captura de datos fue realizado el 8 de octubre de 2016 con un script en el lenguaje de programación R (R Core Team, 1993), recogiendo las respuestas de los servidores de la plataforma en formato JSON (JavaScript Object Notation). Una vez analizadas estas respuestas, se obtuvieron los siguientes datos sobre las 80.459 fotografías descargadas:

- Identificador único de la foto
- Título de la foto
- URLs con la dirección dónde se alojaba la foto
- Coordenadas de latitud y longitud
- Dimensiones de la foto en píxeles

² SightsMap se encuentra disponible en el momento de la redacción en <http://sightsmap.com/>

³ MapSights se encuentra disponible en el momento de la redacción en <https://mapsights.com/>

⁴ Interfaz de Programación de Aplicaciones Web

⁵ La documentación de la Panoramio Data API se encuentra disponible en una captura en Wayback Machine en <http://web.archive.org/web/20160510095912/https://www.panoramio.com/api/data/api.htm> (archivada el 10 de mayo de 2016)

⁶ La documentación de la Panoramio Widget API se encuentra disponible en una captura en Wayback Machine en <http://web.archive.org/web/20160602030004/http://www.panoramio.com/api/widget/api.html> (archivada el 2 de junio de 2016)



- Fecha de ingreso de la foto en el sistema
- Datos sobre la persona usuaria (identificador, nombre y URL de su perfil)

3.2 *Datos institucionales*

El inventario de locales de la ciudad de Barcelona⁷ se descargó de la plataforma del servicio de datos abiertos del Ayuntamiento de Barcelona (OpenDataBCN). Estos datos fueron publicados el 4 de noviembre de 2015 bajo la licencia Creative Commons Attribution 4.0 (CC BY 4.0⁸). El conjunto de datos contenía un total de 78.033 registros, frente a los 67.433 registros del inventario publicado anteriormente, correspondiente al año 2014.

El inventario contiene la localización de estos locales, incluyendo las coordenadas de los locales en el sistema de referencia EPSG:4326 (el mismo que el de los datos de Panoramio), así como en el sistema de referencia ETRS89 / UTM 31N (EPSG:25831), un identificador único asignado por el ayuntamiento, y un conjunto de atributos referentes a las características de estas actividades:

- Información acerca de la actividad del local (activo, sin actividad, o sin información)
- Sector de actividad, clasificado en 8 categorías, de las cuales dos (servicios y comercio) eran las más informativas
- Grupo de actividad, proporcionando información únicamente para las actividades de servicios (9 categorías) y comercio (7 categorías)
- Tipo de actividad, proporcionando información únicamente para las actividades de servicios (23 categorías), comercio (32 categorías) y otras (9 categorías)
- La ubicación del local (frente a calle, en mercado, en galería o en centro comercial)

Adicionalmente, los límites municipales oficiales de la ciudad de Barcelona se obtuvieron del portal de cartografía municipal del Ayuntamiento (CartoBCN⁹), descargando la geometría del término municipal actualizado a fecha 23 de noviembre de 2018. Desarrollo

3.3 *Planteamiento*

Tanto las fotografías geolocalizadas (indicador del interés fotográfico), como la localización de los comercios (indicador de la actividad) en los distintos ámbitos de la ciudad, consistían en patrones de puntos cuyo análisis permite inferir las características del proceso que las ha generado. Existe una amplia literatura acerca del análisis de los patrones espaciales de puntos (Diggle, 2014; Baddeley, Rubak, y Turner, 2016), dónde los puntos son denominados “eventos” en el sentido en que suponen la materialización de un proceso subyacente del cual se quiere obtener un modelo.

Aplicado al caso de estudio, el modelo describiría el atractivo fotográfico de un lugar, que no es uniforme dentro de la ciudad y que conlleva que exista una mayor cantidad relativa de personas que quieren fotografiar cada entorno urbano particular; en algunos casos esta mayor presencia por personas tiene la consecuencia que una de ellas tome una fotografía y la comparta en una

⁷ El inventario de locales de la ciudad de Barcelona se encuentra disponible en el momento de la redacción en <https://opendata-ajuntament.barcelona.cat/data/es/dataset/cens-activitats-comercials>

⁸ Los detalles de la licencia Attribution 4.0 International (CC BY 4.0) se pueden consultar en el momento de la redacción en la web <https://creativecommons.org/licenses/by/4.0/>

⁹ El portal de cartografía del Ayuntamiento de Barcelona se encuentra disponible en el momento de la redacción de manera gratuita (requiere registro) en <http://w20.bcn.cat/cartobcn/>



red social, generando un evento. Sin embargo, este modelo simplificado parte de la base que la única variable que hace variar la cantidad de fotografías es el atractivo fotográfico.

Sin embargo, a igualdad de atractivo fotográfico, cuanto mayor es el flujo peatonal existen mayores posibilidades de que una persona tome una fotografía. En el análisis que nos ocupa se consideró que la distribución espacial de las personas es variable, y por lo tanto debe tenerse en cuenta la variabilidad en el uso del espacio público, incluyéndola como variable explicativa dentro del modelo. De esta manera, se analizó el atractivo teniendo también en consideración distribución espacial de la intensidad de uso dentro de la ciudad, de manera indirecta a partir de los datos de la actividad comercial.

3.4 *Indicador de actividad comercial*

La estimación de la actividad se realizó en base al inventario de locales de la ciudad de Barcelona, considerando que se trataba de un buen indicador de la intensidad de la actividad en los distintos ámbitos de la ciudad, bajo la premisa que una mayor densidad de locales está aparejada a una mayor actividad peatonal. Sin embargo, para que reflejara de manera más fidedigna esta relación, los datos obtenidos fueron filtrados para incluir por su mayor representatividad únicamente los catalogados como comercio o servicios (22.747 y 31.773 registros, respectivamente), que de manera combinada representaban el 70% del total del inventario, excluyendo el resto de usos, así como los locales sin actividad (vacíos o en reforma) y los locales de los cuales no se disponía de información acerca de su actividad.

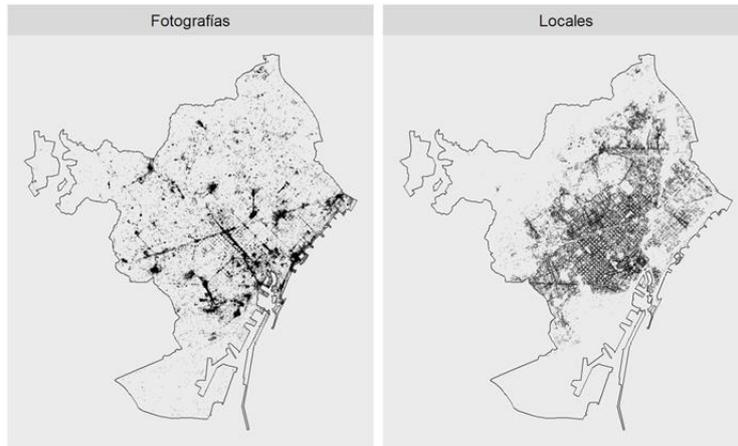
Adicionalmente, debido a la mayor permeabilidad de las plantas inferiores (Valls, Garcia-Almirall, Redondo, y Fonseca, 2014), se mantuvieron únicamente los locales correspondientes a las plantas bajas (identificadas en la base de datos como locales a nivel de calle), excluyendo los locales en mercados, galerías y centros comerciales, tipologías que acostumbran a aislarse de las calles colindantes, creando espacios cerrados en sí mismos a los que se ingresa por unos pocos accesos. Esta consideración descartó un 5% adicional de locales del inventario.

3.5 *Distribución espacial de los patrones de puntos*

Las dos variables presentaban una distribución espacial muy distinta, como se puede apreciar visualmente inspeccionando los mapas de puntos de ambos (Figura 1), donde los puntos correspondientes a las fotografías se concentran en unos ámbitos concretos, mientras que los locales presentan una distribución espacial comparativamente más homogénea.

Existen múltiples metodologías para cuantificar el grado de concentración (*clustering*) de los patrones espaciales de puntos. Por su complejidad, se acostumbran a sintetizar en forma de distribución acumulativa (en función de un radio de búsqueda), en contraste con las distribuciones univariadas que se pueden resumir en una estadística consistente en un único valor numérico. Para cuantificar los dos patrones se calcularon las estadísticas $L(r)$ y $F(r)$ para obtener una caracterización multi-escala de la distribución espacial de ambos patrones de puntos. En ambos casos se aplicó la corrección de borde (Ripley, 1981) para compensar los efectos de la forma de la ventana de observación, y se incluyó una distribución de referencia para la hipótesis nula bajo la cual el patrón de puntos es el resultado de un proceso de Poisson.

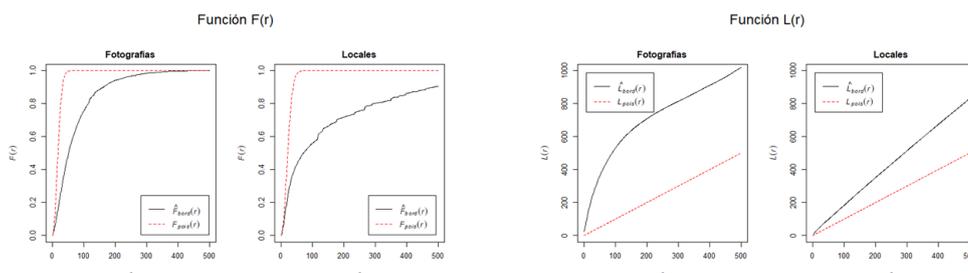
Figura 1. Mapas de los patrones de puntos dentro de la ventana de observación del término municipal para las fotografías de Panoramio (izquierda) y los locales a pie de calle (derecha). Los puntos están grafiados con una opacidad del 10%. Los mapas usan el sistema de referencia EPSG:25831



Fuente: Elaboración propia. Cartografía de los límites municipales de CartoBCN (CC BY 3.0).

La función $F(r)$ es la función de “espacio vacío”, también llamada “distribución de primer contacto esférico” y mide la distribución de la distancia al evento más próximo. Proporciona la probabilidad (en un rango de 0 y 1) de encontrar un punto dentro de un radio r en cualquier punto de la ventana de observación. Como se aprecia en las gráficas (Figura 2, izquierda), con valores pequeños de radio de búsqueda, encontramos muchas más fotografías que negocios. En particular, el 50% y 75% de los puntos se encuentran a 52m y 98m respectivamente para las fotografías, mientras que estos mismos porcentajes se encuentran a 71m y 240m para las localizaciones de los negocios.

Figura 2. Funciones $F(r)$ y $L(r)$ para la distribución espacial de las localizaciones de fotografías y locales. La distribución de referencia de la hipótesis nula (correspondiente a un proceso de Poisson) se muestra en rojo y línea discontinua



Fuente: Elaboración propia.

A pesar de ser menos intuitiva de interpretar por no consistir en una probabilidad, la función $K(r)$ consiste en la media acumulativa del número de puntos dentro de una distancia r de un punto típico, dividido por la intensidad para poder comparar patrones con distinto número de observaciones. Para hacer el gráfico más interpretable, se acostumbra a transformar la distribución de Poisson de referencia en una recta, resultando en la función $L(r)$. Valores por encima de la línea son indicativas de concentración, mientras que por debajo sugieren dispersión. Como se aprecia en las gráficas (Figura 2, derecha), ambas distribuciones espaciales sugieren concentración, pero las fotografías la presentan en un grado mucho más acusado que los locales.



3.6 Estimación de la intensidad

Para poder comparar ambos patrones espaciales de puntos, se calculó la estimación de intensidad para ambos conjuntos de datos, denominada *kernel density estimation* (KDE) en inglés, en contraste con otras aproximaciones que requieren definir unidades de agregación poligonales (Feick y Robertson, 2015).

El resultado del cálculo del KDE es la estimación no paramétrica de la intensidad del proceso que generó hipotéticamente el patrón de puntos, expresado como la expectativa del número de puntos por unidad de área en cada localización, aunque a menudo se confunde con una probabilidad y se usa de manera indistinta como tal. De manera intuitiva, Baddeley, Rubak y Turner (2016, p. 168) describen el proceso de la siguiente manera: “Nuestra analogía favorita es imaginar que se pone una pastilla de chocolate en cada localización. Usando un secador de pelo se aplica calor con lo que el chocolate se funde ligeramente. El resultado es una superficie ondulada de chocolate; la altura de la superficie representa la función de intensidad estimada del proceso.

La masa total de chocolate no varía.”

Para el cálculo se utilizó el paquete *spatstat* (Baddeley y Turner, 2005) que proporciona funcionalidad de análisis de patrones de puntos en el lenguaje de programación R, para conseguir la máxima exactitud¹⁰ en el cálculo. Aunque el resultado es una función continua, para visualizarla es necesario discretizar el resultado, evaluando la función en una matriz de píxeles, y para ello se eligió un tamaño de píxel de 25m de lado, resultando en una matriz de 672x591 puntos, con un total de 163.782 píxeles efectivos dentro de la ventana de observación (41% del total).

La función de núcleo (*kernel*) elegida fue la gaussiana, de manera que el error posicional se distribuyó alrededor de la localización obtenida usando una distribución normal. La desviación estándar de la distribución se fijó en 300m, ligeramente por encima de la regla de Scott (Scott, 1992), que sugiere un valor de 296 para el conjunto de ambos datos, usando un *kernel* Gaussiano isotrópico.

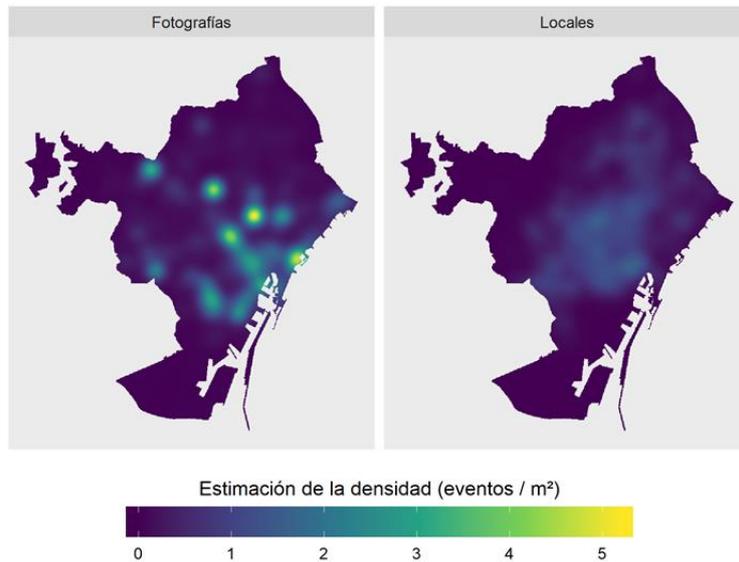
Se aplicó la corrección de borde dentro de la ventana de observación, y para que la suma de la estimación en cada píxel coincidiera exactamente con el número de puntos, se aplicó adicionalmente la corrección de Jones-Diggle (Jones, 1993), forzando además que los valores de estimación de densidad fueran siempre positivos.

Los mapas de intensidad resultantes (Figura 3) se representaron con la escala de color perceptualmente uniforme¹¹ (Garnier, 2017), mostrando la mayor concentración de los puntos correspondientes a las fotografías, en comparación con la distribución espacial de los locales. Los mapas sugieren que algunas zonas de alta intensidad coinciden con una mayor presencia de locales, mientras que en otros espacios de la ciudad esta mayor intensidad no podría ser explicada por un mayor nivel de actividad, evidenciando la necesidad de desarrollar un modelo que contemple la heterogeneidad de la intensidad de uso de la ciudad.

¹⁰ El cálculo utiliza la transformada rápida de Fourier, Fast Fourier Transform (FFT) en inglés

¹¹ Desarrollada para la biblioteca *matplotlib* de Python y llamada “*viridis*”

Figura 3. Mapas de la densidad de puntos dentro de la ventana de observación del término municipal para las fotografías de Panoramio (izquierda) y locales a pie de calle (derecha). Los valores se han convertido en unidades de eventos por m². Los mapas usan el sistema de referencia EPSG:25831



Fuente: Elaboración propia. Cartografía de los límites municipales de CartoBCN (CC BY 3.0).

3.7 Metodología de comparación

El cálculo del KDE permitió convertir los patrones de puntos a comparar (variable y covariable) en sendas funciones, evaluadas en cada uno de los píxeles de una rejilla de puntos, obteniendo valores numéricos para cada una de las celdas de esta rejilla con un valor correspondiente a la estimación del número de eventos por unidad de área, para ambos conjuntos de datos. A través de esta técnica, fue posible comparar las distribuciones espaciales de ambos patrones de puntos, sacrificando la mayor resolución espacial de los datos en bruto, pero obteniendo pares de valores para cada una de las celdas de la rejilla. Sin embargo, para hacer la metodología reproducible, fue necesario encontrar una técnica para comparar los dos conjuntos de valores numéricos de una manera interpretable y robusta.

En primer lugar, fue necesario definir una función para comparar ambos valores. Para facilitar la interpretabilidad de los resultados, se restringió la comparación a operaciones aritméticas básicas de tipo binario, imponiendo que la operación conllevara una noción de direccionalidad (descartando por lo tanto la suma y el producto). Quedando reducidas las opciones a la diferencia y la división, se descartó esta última por la distribución de los valores numéricos de los KDE, que por definición consisten en valores que tienden a cero al aumentar la resolución de la rejilla, con la consiguiente la posibilidad de divisiones entre valores muy cercanos o iguales a cero.

Mientras que la distribución espacial del cociente entre dos valores puede ser de difícil interpretación (en particular con valores inferiores a la unidad), debido a que evalúa la relación entre proporciones, la diferencia mantiene una relación lineal entre los valores, haciendo la lectura de la distribución espacial de la diferencia más intuitiva. Sin embargo, debido a que los valores a comparar provienen de la estimación de la intensidad por KDE, no es posible su



comparación directa debido a que el número de eventos de ambos patrones de puntos no es coincidente.

Para permitir esta comparación se desarrolló un indicador llamado “atractivo relativo” que se define conceptualmente de la siguiente manera (invirtiendo la analogía de Baddeley, Rubak y Turner anteriormente citada): en primer lugar, se supone una condición de partida donde cada uno de los píxeles efectivos (dentro de la ventana de observación) contiene el mismo valor arbitrario, que puede establecerse en un valor unitario, y que puede visualizarse como limaduras de hierro espolvoreadas uniformemente sobre el ámbito de estudio. A continuación, actúa un “campo de fuerza” que tiene una influencia variable en el espacio y que redistribuye estos valores dentro de la ventana de observación, de manera que la suma de todos los valores se mantiene constante (se considera el ámbito de estudio como un sistema cerrado).

3.8 Cálculo

Para realizar el cálculo del atractivo relativo tanto para imágenes como para locales, fue necesario en primer lugar calcular la integral del KDE (dentro de la ventana de observación), correspondiente por definición al área bajo la superficie 3D resultado de su cálculo, y que necesariamente es igual al número de puntos en el patrón de puntos original, una vez aplicadas las correcciones descritas en el apartado 4.4.

En segundo lugar, se calculó la integral parcial para cada pixel, correspondiente al volumen que encierra un paralelepípedo cuya base es el cuadrado correspondiente a cada uno de los píxeles de la rejilla, y cuya altura es el valor calculado mediante el KDE en el píxel correspondiente. Con valores pequeños de píxel (resoluciones altas), la suma de estas integrales parciales es por definición la integral sobre la totalidad del área de estudio. Debido a la proyección, se asumió una medida de píxel de 24,97m en la dirección este-oeste y 24,99m en la dirección norte-sur, resultando en un área por píxel de 623,83m².

Finalmente, se normalizaron los valores para que fueran comparables entre patrones de puntos con diferente número de observaciones y/o distribuciones espaciales, con la fórmula detallada a continuación. Aunque N_{px} podría ser cualquier número elegido arbitrariamente, el valor adoptado, correspondiente al número de píxeles, permite que el resultado se adapte a la resolución de análisis, concordando con la definición detallada en el apartado 4.5.

$$I_p = N_{px} \frac{\iint_{px} KDE_p}{\iint_{dom} KDE_p} = N_{px} \frac{A_{px} KDE_p}{N_p}$$

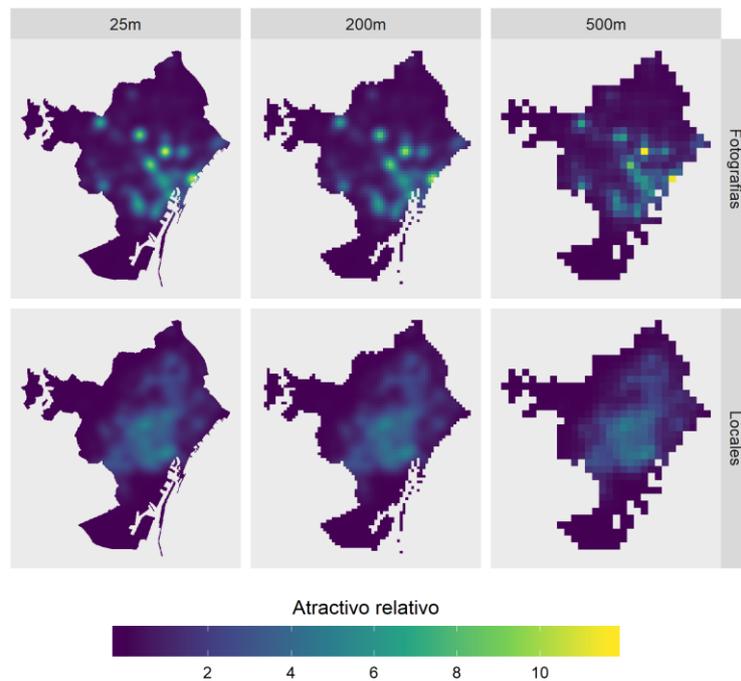
Dónde:

- p** Patrón de puntos para el cual se calcula el indicador (fotografías o locales)
- px** Cada uno de los píxeles del área de estudio (dom), dependiente de la resolución
- dom** Área de estudio (dominio espacial)
- I_p** Indicador de atractivo relativo del patrón de puntos p
- KDE_p** Estimación de densidad (KDE) del patrón de puntos p
- A_{px}** Área del píxel en el sistema de proyección definido
- N_{px}** Numero de píxeles efectivos (dentro del área de estudio)
- N_p** Número de puntos del patrón de puntos p

El diseño de este indicador permitió convertir los valores de estimación de la intensidad (KDE) en un valor robusto. Esta robustez es constatable (Figura 4) si se compara dicho indicador en

una matriz de mapas (Becker, Cleveland, y Shyu, 1996) en múltiples resoluciones (columnas) para los ambos patrones de puntos correspondientes a la variable explicativa y la covariable de control (filas).

Figura 4. Cálculo del indicador propuesto con distintas resoluciones de análisis, desde 25m a 500m de tamaño nominal de pixel, mostrando la robustez de los resultados respecto a los cambios en el tamaño de la rejilla de discretización y la distribución espacial de los patrones de puntos, para el mismo ancho de banda de 300m. Los mapas usan el sistema de referencia EPSG:25831



Fuente: Elaboración propia. Cartografía de los límites municipales de CartoBCN (CC BY 3.0).

Frente a los resultados del KDE, el atractivo relativo presenta las siguientes ventajas: 1) Los valores no son la intensidad por unidad de área, que resulta difícilmente interpretable, sino una medida del atractivo de una celda concreta en relación al conjunto de celdas del ámbito de estudio, que puede considerarse más intuitivo y comparable entre distintos conjuntos de datos, 2) los valores son proporcionales a la intensidad, con un mínimo situado en el cero y una expectativa (hipótesis nula) situada en el valor unitario, siendo la escala lineal (un valor de dos significa que un píxel tiene el doble de interés de lo que sería esperable), 3) el valor obtenido no depende número y/o distribución espacial (*clustering*) de los patrones de puntos a partir de los cuales se computa el KDE, y 4) el resultado no es afectado por la resolución de discretización de la rejilla píxeles ni por la geometría de la ventana de observación (área de estudio), y por lo tanto es invariante respecto a la escala de análisis.

4. Resultados

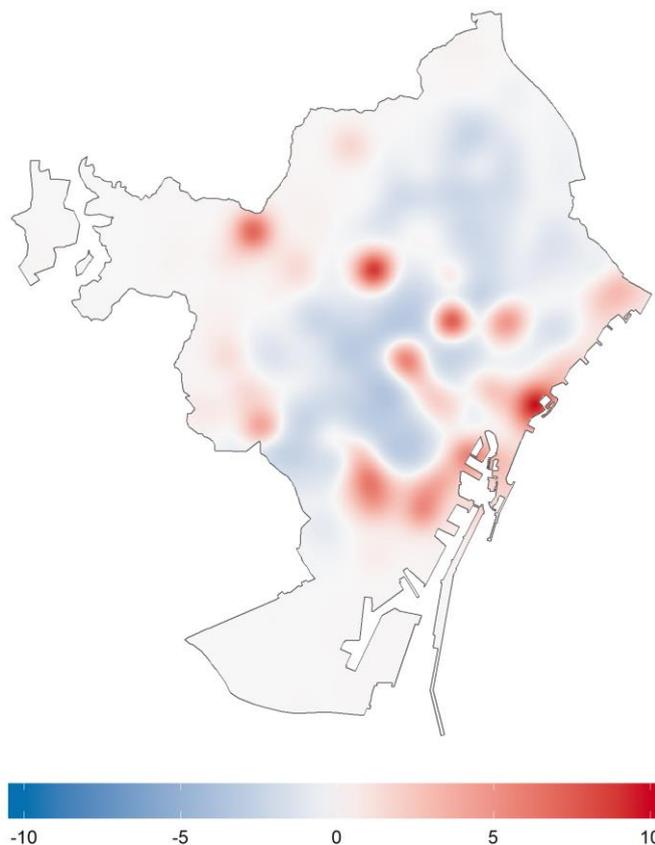
4.1 Atractivo Relativo

El resultado del proceso de normalización descrito convirtió en comparables ambos conjuntos de datos, minimizando los potenciales factores de distorsión descritos en la sección 4.6. Para analizar la intensidad de atractivo fotográfico, controlando la variación en la intensidad del uso

de la ciudad a través de la intensidad de negocios, se calculó la diferencia entre ambos valores para obtener un nuevo indicador, sustrayendo los valores calculados para los negocios a los valores correspondientes a las fotografías.

El resultado puede ser representado en un mapa (Figura 5), con la misma resolución de píxel de la discretización del KDE (25m). En este mapa, la expectativa en cualquier píxel es que la diferencia entre ambos valores sea cero, correspondiente a la hipótesis nula de que el interés fotográfico y la actividad en la ciudad van aparejados. En el mapa, valores positivos (puntos calientes) se muestran en una gama de colores rojos y valores negativos (puntos fríos) se muestran en una gama de colores azules, siguiendo un esquema de color divergente (Harrower y Brewer, 2003). Los dos polos (caliente y frío) utilizan colores más intensos (más oscuros y saturados) a medida que los valores son más extremos, mientras que valores cerca cero, que resultan poco significativos, aparecen en colores cerca del gris (menor saturación).

Figura 5. Mapa ámbitos calientes y fríos, calculado a partir de la diferencia entre la distribución espacial de fotografías (interés fotográfico) y la distribución espacial de locales en planta baja (intensidad de actividad de la ciudad). El rango de valores oscila entre -3,6 y 9,9 pero la escala de color se ha definido de manera simétrica alrededor de cero entre -10 y 10. El mapa usa el sistema de referencia EPSG:25831



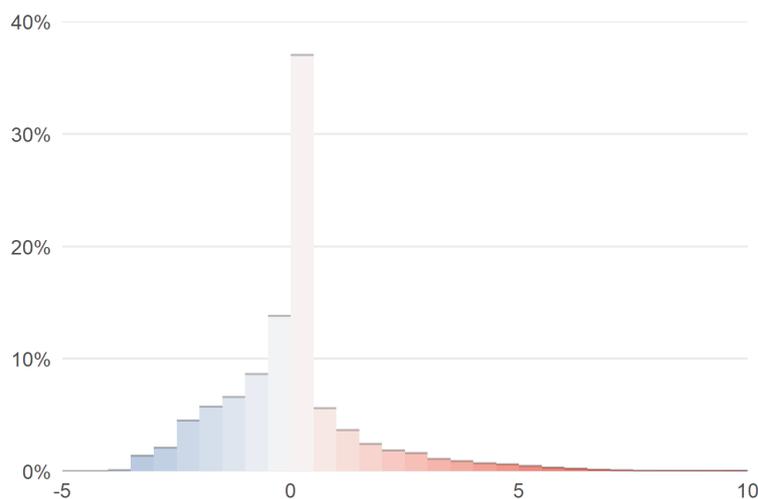
Fuente: Elaboración propia. Cartografía de los límites municipales de CartoBCN (CC BY 3.0).

En el caso que el valor sea superior a cero, se interpreta que existe un interés fotográfico inusualmente alto, relativo a lo que podría explicarse por la presencia más o menos alta de locales. Esta situación puede darse en el caso que 1) exista una intensidad fotográfica muy

elevada, aunque la intensidad de locales sea relativamente estándar, 2) exista una intensidad fotográfica no especialmente destacable, pero que la intensidad de locales sea sensiblemente menor de lo usual. De la misma manera y con la misma lógica, pero con los valores invertidos, se pueden interpretar los valores correspondientes a los puntos fríos.

Examinando el histograma de los valores de cada uno de los 163.782 píxeles del mapa (Figura 6), se aprecia que la mayoría de puntos se concentran cerca del cero, estando el 65% de los valores entre -1 y 1, indicando una desviación relativamente pequeña. Los valores por encima de este rango (rojos) corresponden al 14,5% del total, mientras que los que están debajo (azules) corresponden al 20,5% restante. El 57% de los valores son positivos y el 43% negativos. Sin embargo, los valores extremos se concentran mayoritariamente en los puntos calientes, como se puede apreciar por el sesgo del histograma que presenta una asimetría (*skewness*) superior a uno, con un valor de 1,8.

Figura 6. Histograma de la distribución de valores del mapa de la Figura 5. La escala vertical indica el porcentaje de píxeles que se encuentran representados por cada barra vertical. Las barras utilizan exactamente el mismo patrón de colores que el mapa



Fuente: Elaboración propia.

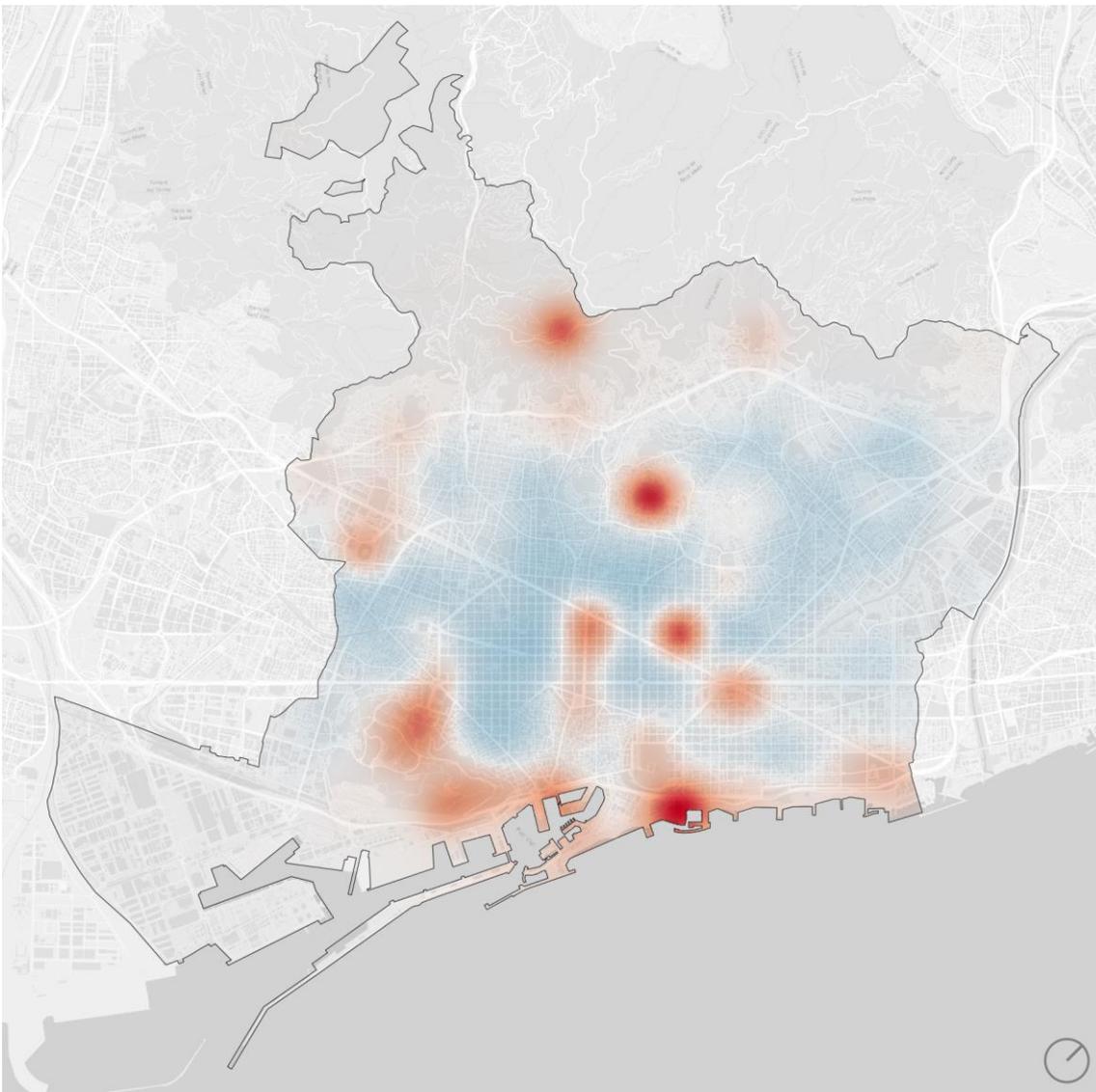
4.2 Discusión

Para visualizar el resultado del cálculo del atractivo relativo en su contexto urbano, se superpuso el ráster obtenido con los valores del indicador a la cartografía base de la ciudad, obtenida del mapa base de ESRI a través de la conexión a su servidor de teselas. La superposición se realizó con el modo “multiply” de QGIS y se convirtió la cartografía base a una escala de grises desaturando completamente la imagen.

Para facilitar la interpretación se rotó el mapa 45° en el sentido de las agujas del reloj, para mostrar el mapa en la orientación convencional en Barcelona, con el mar en la parte inferior y alineado con los ejes principales del Ensanche. Finalmente, la geometría de los límites municipales obtenida de CartoBCN se incluyó como referencia.

El mapa (Figura 7) muestra con claridad las áreas con mayor intensidad fotográfica en la ciudad de Barcelona, pudiéndose distinguir dos tipos de puntos calientes (en rojo): elementos de interés concentrado y ámbitos generales de interés. Por otra parte, también es posible visualizar las zonas con un bajo interés fotográfico (en azul) que podrían denominarse “desiertos”. Finalmente, las áreas de color gris no presentan una divergencia destacable, coincidiendo sus valores con los esperados en base a la actividad estimada a través de la densidad de locales en planta baja.

Figura 7. Resultado del cálculo del indicador superpuesto a la cartografía base y los límites municipales de Barcelona. El mapa usa la proyección WGS 84 / Pseudo-Mercator - Spherical Mercator (EPSG:3857), con una rotación de 45° en el sentido de las agujas del reloj para mostrar la orientación convencional en esta ciudad



Fuente: Elaboración propia. Mapa base de ESRI. Cartografía de los límites municipales de CartoBCN (CC BY 3.0).

Entre los elementos de interés concentrado (puntual) encontramos representada la obra de Antoni Gaudí, pudiéndose apreciar la gran intensidad de la presión fotográfica sobre el Park Güell y la Sagrada Familia. En menor medida, también destacan los puntos de interés correspondientes al ámbito del Parque de atracciones del Tibidabo y la Torre de Collserola, la



Torre Glòries¹² (aunque de manera más difusa debido a que por su tamaño y contexto urbano puede observarse desde una distancia considerable), y el Estadio del Camp Nou. Finalmente, a través a la corrección por intensidad de actividad desarrollado, se pueden apreciar también algunos puntos de interés de menor intensidad, que de otra forma no serían detectados: el Recinto Modernista de Sant Pau, el Laberinto de Horta, el Museo CosmoCaixa, el Monasterio de Pedralbes, los Jardines del Palacio de Pedralbes, el Parque de Cervantes y el Parque de la Ciutadella.

Entre los ámbitos de interés encontramos en primer lugar el litoral, con tres polos identificables: un polo central ubicado en el Puerto Olímpico, y dos polos extremos ubicados respectivamente en el Recinto del Fórum Universal de las Culturas y en el conjunto de la Barceloneta y el Moll de la Fusta.

Otra área claramente apreciable es el eje del Paseo de Gracia y sus calles colindantes hasta adentrarse en Las Ramblas, que destaca por su atractivo a pesar del gran número de locales presentes, con un polo dominante situado en La Pedrera. Finalmente, destacan dos áreas de interés diferenciadas en Montjuïc: en el lado mar con el conjunto del Castillo de Montjuïc y el Jardín Botánico, y en el lado montaña entre la Fuente Mágica y el Palacete Albéniz.

5. Conclusiones

La metodología desarrollada permite identificar de manera robusta los puntos calientes de interés fotográfico, controlando la importante heterogeneidad del nivel de actividad dentro de la ciudad. La incorporación de esta covariable, a través de la intensidad de locales, permite identificar áreas donde las personas acuden únicamente por su interés fotográfico, que de otra manera podrían resultar desapercibidas.

La distribución de estos puntos de interés refleja el interés por la obra de Antoni Gaudí, al mismo tiempo que revela grandes ámbitos de la ciudad donde los turistas raramente se adentran (desiertos). El mapa resultante revela la escasa conectividad entre los puntos de interés, que aparecen claramente aislados, información que debería permitir optimizar el diseño de la movilidad turística dentro de la ciudad, fomentando por ejemplo el transporte público entre estos puntos, o bien impulsando una red de conectividad peatonal que incluya puntos de interés intermedios.

5.1 Futuras líneas de trabajo

Los resultados para Ciutat Vella sugieren que, aunque su atractivo es alto, es parcialmente compensado por un alto número de comercios y servicios. Sin embargo, este resultado no refleja completamente la realidad, debido a que muchos locales tienen una actividad directa y exclusivamente relacionada con el turismo. Se prevé ajustar la metodología para minimizar este posible sesgo en futuras investigaciones, a través del análisis de los residuos en un modelo de regresión espacialmente autorregresivo.

Conflicto de Intereses: Los autores declaran que no hay conflicto de intereses.

¹² También conocida como Torre Agbar



Bibliografía

- Anselin, L., y Williams, S. (2015). Digital neighborhoods. *Journal of Urbanism: International Research on Placemaking and Urban Sustainability*, 9(4), 305-328. DOI <https://doi.org/10.1080/17549175.2015.1080752>
- Baddeley, A., Rubak, E., y Turner, R. (2016). *Spatial point patterns: Methodology and applications with R*. Boca Raton; London; New York: CRC Press, Taylor y Francis Group.
- Baddeley, A., y Turner, R. (2005). spatstat: An R Package for Analyzing Spatial Point Patterns. *Journal of Statistical Software*, 12(6). DOI <https://doi.org/10.18637/jss.v012.i06>
- Becker, R. A., Cleveland, W. S., y Shyu, M.-J. (1996). The Visual Design and Control of Trellis Display. *Journal of Computational and Graphical Statistics*, 5(2), 123-155. DOI <https://doi.org/10.1080/10618600.1996.10474701>
- Diggle, P. (2014). *Statistical analysis of spatial and spatio-temporal point patterns* (tercera edición). Boca Raton: CRC Press, Taylor y Francis Group.
- Feick, R., y Robertson, C. (2015). A multi-scale approach to exploring urban places in geotagged photographs. *Computers, Environment and Urban Systems*, 53, 96-109. DOI <https://doi.org/10.1016/j.compenurbsys.2013.11.006>
- García-Hernández, M., De la Calle-Vaquero, M., y Yubero, C. (2017). Cultural Heritage and Urban Tourism: Historic City Centres under Pressure. *Sustainability*, 9(8), 1346. DOI <https://doi.org/10.3390/su9081346>
- García-Palomares, J. C., Gutiérrez, J., y Mínguez, C. (2015). Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS. *Applied Geography*, 63, 408-417. DOI <https://doi.org/10.1016/j.apgeog.2015.08.002>
- Garnier, S. (2017). *viridis: Default Color Maps from «matplotlib»*. Recuperado de <https://CRAN.R-project.org/package=viridis>
- Goodchild, M. F. (2016). GIS in the Era of Big Data. *Cybergeog: European Journal of Geography*. Recuperado de <https://cybergeog.revues.org/27647>
- Harrower, M., y Brewer, C. A. (2003). ColorBrewer.org: An Online Tool for Selecting Colour Schemes for Maps. *The Cartographic Journal*, 40(1), 27-37. DOI <https://doi.org/10.1179/000870403235002042>
- Jones, M. C. (1993). Simple boundary correction for kernel density estimation. *Statistics and Computing*, 3(3), 135-146. DOI <https://doi.org/10.1007/BF00147776>
- Mansilla, J. A., y Milano, C. (2019). Becoming centre: Tourism placemaking and space production in two neighborhoods in Barcelona. *Tourism Geographies*, 0(0), 1-22. DOI <https://doi.org/10.1080/14616688.2019.1571097>
- Peng, R. D. (2011). Reproducible Research in Computational Science. *Science*, 334(6060), 1226. DOI <https://doi.org/10.1126/science.1213847>
- R Core Team. (1993). *R: A Language and Environment for Statistical Computing*. Recuperado de <http://www.R-project.org/>



Ripley, B. D. (1981). *Spatial statistics*. New York, NY: Wiley.

Salas-Olmedo, M. H., Moya-Gómez, B., García-Palomares, J. C., y Gutiérrez, J. (2018). Tourists' digital footprint in cities: Comparing Big Data sources. *Tourism Management*, 66, 13-25. DOI <https://doi.org/10.1016/j.tourman.2017.11.001>

Scott, D. W. (1992). *Multivariate density estimation: Theory, practice, and visualization*. New York: Wiley.

Shoval, N., y Ahas, R. (2016). The use of tracking technologies in tourism research: The first decade. *Tourism Geographies*, 18(5), 587-606. DOI <https://doi.org/10.1080/14616688.2016.1214977>

Valls, F. (2019). *Digital Traces and Urban Research: Barcelona Through Social Media Data* (Tesis Doctoral, Universitat Politècnica de Catalunya). Recuperado de <http://hdl.handle.net/10803/667168>

Valls, F., Garcia-Almirall, P., Redondo, E., y Fonseca, D. (2014). From Raw Data to Meaningful Information: A Representational Approach to Cadastral Databases in Relation to Urban Planning. *Future Internet*, 6(4), 612-639. DOI <https://doi.org/10.3390/fi6040612>