



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



ESCUELA TÉCNICA SUPERIOR
DE INGENIERÍA GEODÉSICA
CARTOGRÁFICA Y TOPOGRÁFICA
UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



GRADO EN INGENIERÍA GEOMÁTICA Y TOPOGRAFÍA
TRABAJO FIN DE GRADO

Autor: Marcel Buenaventura Fenech

Tutor: Laura Sebastián Tarín

Cotutor: Eliseo Jorge Marzal Calatayud

Septiembre de 2016

Resumen

El principal objetivo de este trabajo es analizar el comportamiento de los turistas en la ciudad de Valencia, partiendo de los datos extraídos de las fotografías de su perfil de la red Flickr. Los datos se almacenan en una base de datos relacional para que su tratamiento y manejo sea más efectivo.

Es esencial distinguir los usuarios que son turistas de los que no. Para esta tarea se seleccionan una serie de criterios lógicos. A partir de estos datos, se calculan parámetros básicos tanto para todos los usuarios como para los turistas. Se crean gráficas en las que se ven las horas de mayor afluencia de turistas, los meses que crece el número de turistas, los puntos de interés más visitados, etc. También se crean planos automatizados de la densidad de fotografías y videos en distintas zonas de la ciudad.

Con estos análisis se demuestra que se puede obtener información de gran interés a partir de la interacción de los usuarios con las redes sociales, y realizar estudios para mejorar el turismo en una ciudad concreta.

Abstract

The main purpose of this Project is to analyze the tourists behavior in the city of Valencia, all this, extrapolating the data of their Flickr' network. The data is stored in a relational data base, which makes its treatment easier and increases the effectiveness of its use.

It is necessary to distinguish the users between the ones who are tourists and the ones who are not, I created a few logical criteria to do so. Through the data it is possible to establish some basic parameters for both, regular users and tourists.

I created graphics where it is possible to appreciate the times of the day when the higher number of tourists was there, the months when the visitants number increases, the most visited points... I also created automatized plans with the density of the pictures and videos of each part of the town.

With this analysis I can prove that it is possible to obtain interesting information from the interaction of the people on the social network and also to realize studies in order to improve the tourism in a specific city.

Agradecimientos

En primer lugar, quiero agradecer a mi tutora, Laura Sebastiá Tarín, la confianza que ha depositado en mí para la realización de este estudio. También agradecer a mi cotutor Eliseo Jorge Marzal Calatayud, haberme transferido los datos de las fotografías.

A mis amigos de la universidad, sin los que el tiempo dedicado al proyecto no hubiese sido tan ameno. Gracias a Macarena Parrizas Siles y a Miguel Lorenzo Perales Solsona por las pausas café y los momentos de humor.

Gracias a mis padres por haber confiado en mí para conseguir un título universitario y apoyarme durante todos estos años.

Por último, aunque por ello no menos importante, muchas gracias a mi compañera sentimental por calmarme cuando estaba agobiado, motivarme cuando estaba vago y aguantarme cuando estaba enfadado y frustrado.

ÍNDICE GENERAL

1.	INTRODUCCIÓN.....	10
1.1.	Introducción	10
1.2.	Objetivo.....	10
1.3.	Motivación.....	11
1.4.	Estructura de la memoria	12
2.	ESTADO DEL ARTE	13
2.1.	Sistema de Información Geográfica (SIG).....	13
2.2.	Big Data	15
2.3.	Flickr	17
3.	TECNOLOGÍAS Y HERRAMIENTAS UTILIZADAS.....	20
3.1.	Tecnologías	20
3.1.1.	PostgreSQL	20
3.1.2.	PostGIS	21
3.1.3.	Python.....	22
3.2.	Herramientas	24
3.2.1.	ArcGis.....	24
3.2.2.	PgAdmin III	25
3.2.3.	IDLE (Python GUI).....	26
3.2.4.	Dia	26
4.	BASE DE DATOS.....	28
4.1.	Fuente de los datos y datos de partida.....	28
4.1.1.	Datos de las fotografías	28
4.1.2.	Cartografía de las ciudades	31
4.1.3.	Cartografía de los puntos de interés	33
4.2.	Diseño de la base de datos.....	37
4.2.1.	Diseño conceptual	37
4.2.2.	Diseño lógico	40
4.2.3.	Diseño físico	41
4.3.	Conclusiones de la base de datos.....	45
5.	APLICACIÓN.....	47

5.1.	Aspecto de la aplicación.....	47
5.2.	Menús de la aplicación.....	48
5.3.	Esquema de la aplicación	49
5.4.	Módulos y librerías	50
6.	ANÁLISIS DE LOS DATOS	53
6.1.	Datos iniciales.....	53
6.1.1.	Número de usuarios totales	54
6.1.2.	Número de fotografías por ciudad.....	54
6.2.	Análisis general.....	56
6.2.1.	Número de usuarios únicos en Valencia	56
6.2.2.	Número de fotografías por usuario.....	56
6.2.3.	Intervalo de captura de fotografías.....	59
6.2.4.	Intensidad de captura de fotografías	60
6.2.5.	Análisis espacial	64
6.2.6.	Análisis espacio-temporal	65
6.2.7.	Análisis gráfico.....	68
6.3.	Análisis de turistas	69
6.3.1.	Determinación de los usuarios turistas.....	69
6.3.2.	Número de turistas únicos.....	75
6.3.3.	Número de fotografías por usuario turista	75
6.3.4.	Intervalo de captura de fotografías de turistas.....	76
6.3.5.	Intensidad de captura de fotografías de turistas.....	77
6.3.6.	Análisis espacial de turistas	80
6.3.7.	Análisis turista espacio-temporal.....	81
6.3.8.	Análisis gráfico.....	82
6.4.	Comparación de resultados	84
7.	CONCLUSIONES.....	88
7.1.	Conclusiones	88
7.2.	Trabajo futuro.....	89
8.	Referencias bibliográficas.....	90
ANEXO I – Planos de usuarios.....		92
1.	Plano de densidad de fotografías/videos de usuarios en Valencia. Casco histórico y alrededores.....	93

2. Plano de densidad de fotografías/videos de usuarios en Valencia. Ciudad de las Artes y las Ciencias.....	94
3. Plano de densidad de fotografías/videos de usuarios en valencia. Puerto deportivo y playa de la Malvarrosa.	95
4. Plano de densidad de fotografías/videos de usuarios en valencia. Parque de Cabecera y Feria de Valencia (Benimámet).	96
ANEXO II - Planos de turistas	97
1. Plano de densidad de fotografías/videos de turistas en Valencia. Casco histórico y alrededores.....	98
2. Plano de densidad de fotografías/videos de turistas en Valencia. Ciudad de las Artes y las Ciencias.....	99
3. Plano de densidad de fotografías/videos de turistas en valencia. Puerto deportivo y playa de la Malvarrosa.	100
4. Plano de densidad de fotografías/videos de turistas en valencia. Parque de Cabecera y Feria de Valencia (Benimámet).	101

ÍNDICE DE ILUSTRACIONES

Ilustración 1: Componentes de un Sistema de Información Geográfica	15
Ilustración 2: Las 5 V's del Big Data	16
Ilustración 3: Página web principal de Flickr	17
Ilustración 4: Página web de la API de Flickr	18
Ilustración 5: Cabecera de la página web de PostgreSQL	20
Ilustración 6: Icono de la extensión PostGis	21
Ilustración 7: Icono del lenguaje de programación Python	23
Ilustración 8: Tipos de plataformas ArcGis	24
Ilustración 9: Pantalla de carga de PgAdmin III	25
Ilustración 10: Editor SQL con consulta de ejemplo	25
Ilustración 11: Ventana principal IDLE	26
Ilustración 12: Dia. Editor de diagramas	26
Ilustración 13: Tabla multimedia con todos sus campos	30
Ilustración 14: Selección de una geometría OSM	32
Ilustración 15: Búsqueda de una geometría OSM	32
Ilustración 16: Descarga de una geometría OSM	33
Ilustración 17: Tabla ciudades en PgAdmin III	33
Ilustración 18: Web del Portal de Transparencia y Datos Abiertos.....	33
Ilustración 19: Catálogo de datos abiertos disponibles	34
Ilustración 20: Conjunto de datos de Monumentos turísticos.....	34
Ilustración 21: Lista de monumentos y sus coordenadas.....	35
Ilustración 22: Conexión con el segundo programa	36
Ilustración 23: URL de Catastro con coordenadas.....	36
Ilustración 24: URL de Catastro con la referencia catastral	36
Ilustración 25: Tabla de monumentos en PgAdmin III.....	36
Ilustración 26: Entidad del diseño conceptual	38
Ilustración 27: Atributo del diseño conceptual.....	39
Ilustración 28: Relación del diseño conceptual	39
Ilustración 29: Diagrama Entidad-Relación del esquema conceptual.....	40
Ilustración 30: Sentencia SQL para la tabla multimedia.....	43
Ilustración 31: Sentencia SQL para la tabla city.....	44
Ilustración 32: Sentencia SQL para la tabla punto_interes.....	44
Ilustración 33: Atributo de tipo punto con EPSG 25830	46
Ilustración 34: Atributo de tipo poligonal con EPSG 25830	46
Ilustración 35: Sentencia de transformación de sistema de coordenadas.....	46

Ilustración 36: Ventana principal de la aplicación	47
Ilustración 37: Icono de la aplicación	48
Ilustración 38: Esquema de la aplicación.....	49
Ilustración 39: Captura de los módulos y librerías utilizadas.....	50
Ilustración 40: Logo Psycopg.....	51
Ilustración 41: Logo oficial de Numpy	51
Ilustración 42: Logo oficial de Matplotlib	51
Ilustración 43: Pantalla de número de usuarios totales.....	54
Ilustración 44: Gráfica de número de fotografías por ciudad.....	55
Ilustración 45: Pantalla de número de usuarios en Valencia.....	56
Ilustración 46: Gráfica fotografías por usuario	57
Ilustración 47: Gráfica fotografías por usuarios divididos en dos grupos.....	58
Ilustración 48: Pantalla de intervalo de captura	60
Ilustración 49: Gráfica fotografías por año de captura	61
Ilustración 50: Gráfica fotografías por mes de captura.....	62
Ilustración 51: Gráfica fotografías por día de captura	63
Ilustración 52: Gráfica fotografías por hora de captura	64
Ilustración 53: Gráfica fotografías por punto de interés.....	65
Ilustración 54: Gráfica fotografías por día de captura en POIs.....	66
Ilustración 55: Gráfica fotografías por día de captura en Pois.....	67
Ilustración 56: Miniatura del plano de densidad de fotografías de usuario	68
Ilustración 57: Gráfica fotografías por usuario turista.....	76
Ilustración 58: Pantalla de intervalo de captura de turistas	77
Ilustración 59: Gráfica fotografías turistas por año de captura	77
Ilustración 60: Gráfica fotografías turistas por mes de captura	78
Ilustración 61: Gráfica fotografías turistas por día de captura	79
Ilustración 62: Gráfica fotografías turistas por hora de captura	79
Ilustración 63: Gráfica fotografías turistas por puntos de interés.....	80
Ilustración 64: Gráfica fotografías turistas por días en POIs	81
Ilustración 65: Gráfica fotografías turistas por hora en POIs.....	82
Ilustración 66: Miniatura del plano de densidad de fotografías de turistas	83
Ilustración 67: Ofertas de la web Turisvalencia.....	84
Ilustración 69: Representación de fotografías de Flickr. Casco histórico	85
Ilustración 70: Representación de fotografías de Flickr. Artes y Ciencias	85
Ilustración 71: Representación de fotografías de Flickr. Puerto y playa	86
Ilustración 72: Representación de fotografías de Flickr. Parque cabecera.....	87

ÍNDICE DE TABLAS

Tabla 1: Clave ajena de la entidad multimedia	41
Tabla 2: Atributos de la entidad multimedia	42
Tabla 3: Atributos de la entidad city	43
Tabla 4: Atributos de la entidad punto_interes.....	44

1. INTRODUCCIÓN

En esta parte de la memoria, daré una información general del proyecto. Expondré el objetivo principal, la motivación por la cual he optado por este tema y por último, una breve descripción de los apartados de la memoria para dar una visión generalizada de los apartados que se van a tratar.

1.1. Introducción

La memoria forma parte de la última etapa del Grado en Ingeniería en Geomática y Topografía, para ser más específicos, de una asignatura llamada “Trabajo fin de Grado”.

El objetivo principal de este trabajo, es que el alumno recicle todos los conocimientos obtenidos a lo largo de los años de estudio del Grado, conectando estos conocimientos para que formen un todo. Además, en el proyecto se emplean tecnologías nuevas, que no han sido estudiadas durante el grado, por lo que se demuestra la capacidad de un aprendizaje autónomo, y la obtención de nuevas habilidades que complementan lo estudiado durante el grado.

Por otra parte, se pretende demostrar si el alumno es capaz de desenvolverse en un terreno más hostil, es decir, en un proyecto que no se encuentre adulterado para que exista un determinado resultado, y de esta manera debe encontrar distintos caminos para su resolución.

En resumen, la intención del trabajo fin de grado es que el alumno mezcle los conocimientos del grado con conocimientos autodidácticos, demostrando su capacidad para adquirirlos y, además, sus habilidades para desenvolverse fuera de un marco educativo preestablecido. Por primera vez a lo largo de sus estudios, debe enfrentar un trabajo similar a los que podría encontrar en el mundo laboral, y por ello, consigue demostrar su capacidad para enfrentarse al siguiente paso de su educación.

1.2. Objetivo

El objetivo principal del trabajo es realizar un análisis de los datos almacenados por Flickr (yahoo.com). Esta plataforma contiene una gran cantidad de fotografías y videos tomadas por todo el mundo.

Para proceder con el análisis, he definido los siguientes objetivos:

- Creación de una base de datos para albergar toda la información referida a las fotografías y videos descargados, geometría de las ciudades y geometría de los puntos de interés turístico (POIs¹).
- Determinación de los usuarios turistas mediante una serie de criterios lógicos.
- Desarrollo de una aplicación para la manipulación y la visualización de todos los datos.
- Realizar un análisis general (todos los usuarios de la ciudad), un análisis de turistas (usuarios que se consideran turistas) y un análisis gráfico (mapas de densidad).
- Creación de distintas gráficas y mapas a partir a partir de los datos descargados.
- Interpretación de los resultados obtenidos en los análisis.

1.3. Motivación

Por todos es sabido que España es un destino turístico apreciado por todos. En los últimos años ha habido un aumento progresivo de turistas que llegan a nuestra región. Por poner unos ejemplos, España recibió 60.6 millones de turistas en el año 2013, un 5.6% más que en el año anterior, esta cifra no se alcanzada desde hacía muchos años. En el año 2014 se vuelve a superar la marca del año pasado, estableciendo un total de 65 millones de turistas, un 7.1% más que el anterior. Por último, en el año 2015, vuelve a subir el número de turistas hasta superar los 68 millones. Viendo estos datos no cabe duda de que España tiene que aprovechar su interés turístico.

Actualmente, estos datos sobre el turismo los obtiene el INE (Instituto Nacional de Estadística) por medio de una serie de encuestas: Encuesta de Ocupación Turística (OCUPATUR), Encuesta de Turismo de Residentes (ETR/FAMILITUR) y Estadística de Movimientos Turísticos en Fronteras (FRONTUR). Se las realizan directamente a los propios turistas y a los establecimientos turísticos (hoteles, atracciones, etc.).

Entre las comunidades de destino turístico más populares se encuentra Catalunya, Madrid, Islas Baleares, Andalucía y la Comunidad Valenciana.

Con este trabajo demostraré que es posible obtener información complementaria a las encuestas a partir del análisis de la interacción de los usuarios con las redes sociales. Permitiendo mejorar el turismo en la ciudad gracias a la información extraída.

¹ Point of Interest.

1.4. Estructura de la memoria

He dividido la memoria en diversos apartados, que quedan definidos de la siguiente manera:

1. **Introducción**
2. **Estado del Arte**
3. **Tecnologías y herramientas utilizadas**
4. **Base de datos**
5. **Aplicación**
6. **Análisis de los datos**
7. **Conclusión**

A lo largo del trabajo describiré el objetivo primordial de cada uno de estos capítulos y los analizaré en profundidad mediante diversos apartados.

2. ESTADO DEL ARTE

En este capítulo se describe la base teórica sobre la cual se sustenta todo el contenido del proyecto. Para ello he definido dos conceptos que considero primordiales para la ejecución del proyecto, los Sistemas de Información Geográfica (SIG) y el Big Data. Finalmente explicaré la aplicación web (Flickr) que he utilizado como fuente de datos para la realización de los análisis.

2.1. Sistema de Información Geográfica (SIG)

Los Sistemas de Información Geográfica (SIG o GIS por sus siglas en inglés), han sido definidos de diversas maneras por multitud de autores. A continuación cito las definiciones que considero más significativas [1]:

- “Conjunto de herramientas para seleccionar, almacenar, recuperar, transformar y exhibir datos espaciales del mundo real para un sistema particular con propósito definido” (*Burrough, 1986*).
- “Un sistema computarizado que proporciona cuatro sistemas de capacidades para manejar datos georreferenciados: entrada de datos, administración de datos (almacenaje y recuperación de datos) manipulación, análisis y salida de datos” (*Arnoff, 1989*).
- “Conjunto de hardware, software y procedimientos elaborados para facilitar la obtención, gestión, manipulación, análisis, modelado, representación y salida de datos espacialmente referenciados, para resolver problemas complejos de planificación y gestión” (*National Center for Geographic Information & Analysis, 1990*).

Otra definición que considero más actualizada y precisa es la siguiente: “integración organizada de hardware, software, datos geográficos y personal, diseñada para capturar, almacenar, manejar, analizar, modelar y representar en todas sus formas la información geográficamente referenciada con el fin de resolver problemas complejos de planificación y gestión”. También puede definirse como un modelo de una parte de la realidad referido a un sistema de coordenadas terrestre y construido para satisfacer unas necesidades concretas de información.

El SIG funciona como una base de datos con información geográfica (datos alfanuméricos) que se encuentra asociada por un identificador común a los objetos gráficos de un mapa digital. De esta forma, señalando un objeto se conocen sus

atributos y, de forma inversa, preguntando por un registro de la base de datos se puede saber su localización en la cartografía.

La razón fundamental para utilizar un SIG es la gestión de información espacial. El sistema permite separar la información en diferentes capas temáticas y las almacena independientemente, permitiendo trabajar con ellas de manera rápida y sencilla. Es decir que facilita al usuario la posibilidad de relacionar la información existente a través de la topología de los objetos, con el fin de generar otra nueva que no podríamos obtener de otra forma.

Un SIG está formado por cinco componentes o elementos, cada uno cumple con una función específica, que permite que interaccionen entre ellos. Es decir, recoge la información para que sea procesada o se realice un tratamiento, de manera que los recursos técnicos, humanos y las metodologías forman un único ente. La descripción de los componentes es la siguiente:

1. **Hardware:** es el equipo físico con el que opera un SIG. Actualmente el software de estos sistemas se ha adaptado a diversos tipos de hardware desde arquitecturas clientes-servidor hasta computadoras de escritorio aisladas. Para las consultas espaciales, el hardware es útil para efectuar el procesamiento de las operaciones que, con base a algoritmos, solucionan las relaciones entre geometrías.
2. **Software:** proporciona las herramientas y funciones necesarias para almacenar, analizar y desplegar la información geográfica.
3. **Datos:** se refiere al elemento principal para lograr una correcta información. Es decir una vez conocido el objeto del modelo del mundo real, se identifican las propiedades que lo forman, por ejemplo, sus atributos que se refieren a los elementos descriptivos y el tipo de geometría como el elemento espacial. En las consultas espaciales es necesario conocer el tipo de geometría entre los objetos del mundo real que se relacionan topológicamente.
4. **Personal:** son las personas que se encargan de administrar el sistema así como de desarrollar un proyecto basado en el mundo real, entre los que se involucran analistas, desarrolladores, administradores, programadores, y usuarios.
5. **Métodos:** son los planes de un buen diseño y las normas por parte de la empresa, las cuales son modelos y prácticas de operación que cada organización emplea para la resolución de los problemas de una manera efectiva.



Ilustración 1: Componentes de un Sistema de Información Geográfica

Todos estos componentes tienen la finalidad de establecer la estructura de un SIG e implementar aplicaciones que apoyen la toma de decisiones, como por ejemplo las consultas espaciales, que aunque por sí solas no muestran la solución, permiten que la persona responsable las interprete y pueda tomar una decisión.

2.2. Big Data

Es muy común que cuando se habla de Big Data se haga referencia a grandes cantidades de datos. Pero este concepto es mucho más que eso. En realidad, cuando se habla del Big Data, se refiere a la gestión y análisis de enormes volúmenes de datos que no se podrían tratar de una manera tradicional, debido a que superan los límites y capacidades de las herramientas de software y de hardware utilizadas habitualmente.

En los últimos años, los usuarios han cambiado la manera de interactuar en la red, las revoluciones tecnológicas como la web 2.0 (foros, redes sociales, multimedia...) y dispositivos móviles facilitan a los usuarios la conectividad constante y, por tanto, la generación de grandes cantidades de información, que puede ser interpretada y analizada para obtener un beneficio empresarial.

El concepto de Big Data que vamos a utilizar se caracteriza por el empleo del modelo de las 5V's que son las siguientes:

- **Volumen:** capta y organiza toda la información que se genera a través de la red.

- **Velocidad:** es tiempo tanto para la generación de la información como el de su análisis, y más si se necesitan conclusiones a tiempo real.
- **Variedad:** es la uniformidad de toda la información, es necesario que el origen de los datos sea lo más heterogéneo posible, permitiendo que se pueda combinar cada tipo de dato.
- **Veracidad:** es calidad del dato, se deben encontrar métodos para comprobar la calidad de la información recibida.
- **Valor:** el objetivo del *Big Data* es dotar de valor a la sociedad, empresas, gobiernos y en definitiva a todas las personas para poder ayudar en su desarrollo.



Ilustración 2: Las 5 V's del Big Data

Los datos también se pueden clasificar según su naturaleza o su origen, lo que ayuda a entender mejor la evolución de la explotación de los datos. Existen tres tipos de datos:

- **Datos estructurados:** se trata de la información ya procesada, filtrada y con una estructura clara. Es el tipo de dato que más se utiliza, ya que es la gente suele buscar. Nutren a las bases de datos relacionales.

- **Datos semi-estructurados:** es la información que ya está procesada pero no tiene una estructura. Un ejemplo claro podrían ser los ficheros con información en un lenguaje de etiquetas (HTML, XML, JSON...).
- **Datos no estructurados:** es la información que se encuentra sin procesar y no tiene ninguna estructura. Por ejemplo la información colgada en las redes sociales.

Como he comentado anteriormente la forma más habitual de trabajar hoy en día es mediante datos estructurados, seguida del uso de datos semi-estructurados. Para la elaboración de este proyecto he utilizado los dos tipos de datos, concretamente los datos estructurados empleando la base de datos relacional de PostgreSQL, y datos semi-estructurados obtenidos en formato JSON devuelto por el servidor de la API (Application Programming Interface) de Flickr.

2.3. Flickr

Es un sitio web que permite al usuario almacenar, ordenar, buscar, vender y compartir fotografías y videos en línea. La única condición y requisito para hacerlo es que tanto las fotografías como los videos sean creados por los propios usuarios.

La comunidad de usuarios se rige por una serie de normas de comportamiento y unas condiciones de uso, que favorecen una buena gestión del contenido y una armonía entre los usuarios.

Para poder publicar imágenes y videos es necesario registrarse en la plataforma, lo que puede hacerse de manera gratuita, aunque con limitaciones en cuanto a la capacidad de almacenamiento, con la opción de pago estas limitaciones desaparecen, esta opción recibe el nombre de cuenta “PRO”.

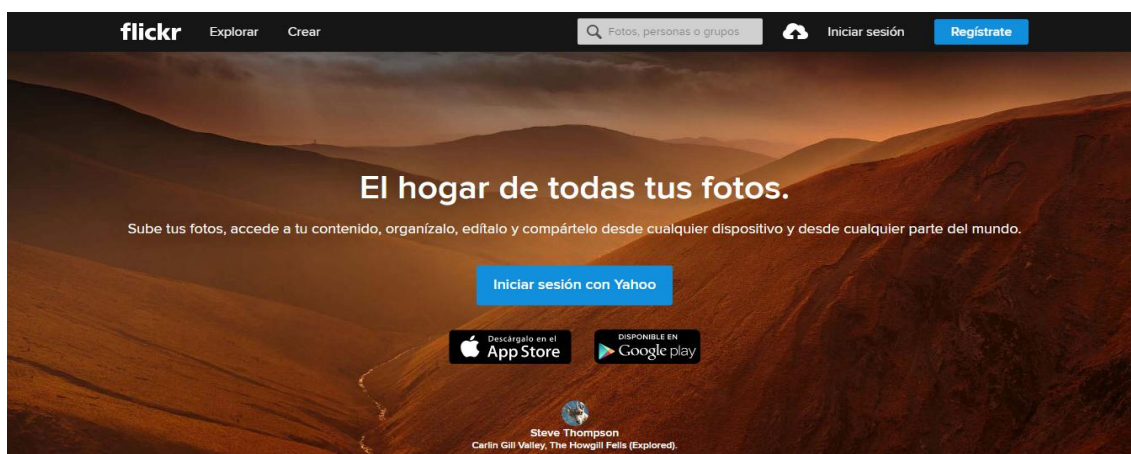


Ilustración 3: Página web principal de Flickr

Flickr se considera una red social, por lo que permite a todos los usuarios comentar y votar las imágenes publicadas por otros usuarios, y crear conversaciones y debates entre ellos. Al interactuar con las fotografías y videos, se abre una página en la que, además de la visualización, permite realizar otro tipo de acciones como escribir comentarios, opiniones, marcarlo como favorito, compartirlo, etc...

También existen los llamados grupos, en los cuales un número de usuarios se junta para compartir contenido de un mismo tipo, es decir contenido de interés común para todos ellos. La gente puede etiquetar sus imágenes y videos, estas etiquetas sirven para catalogar las imágenes sobre un contenido y actúan como palabras clave, ya que muchos de los usuarios intentan que sus archivos lleguen al mayor número de personas posibles. Por eso, estas etiquetas suelen usar como idioma predominante el inglés ya que, es más conocido internacionalmente.

Otra de las características, es que se puede almacenar la ubicación de las imágenes, de esta manera se puede saber en qué lugar del mundo se captura una imagen en concreto.

Por último, hay que comentar que ofrece otras funcionalidades como puede ser la API, la cual permite a desarrolladores independientes de la plataforma crear servicios y aplicaciones propias vinculadas a Flickr.




Ilustración 4: Página web de la API de Flickr

Como podemos ver en la página web de la API, la información se estructura en tres partes. La primera de ellas es una cabecera idéntica a la propia web de Flickr. La segunda de ellas, la columna izquierda, trata la información más relevante de la API, así como aspectos que los desarrolladores deben de tener en cuenta a la hora de crear

aplicaciones. La tercera y última, una columna en la parte derecha, contiene todos y cada uno de los métodos que se pueden emplear para la extracción de información, al acceder a ellos tenemos su explicación y los parámetros de entrada necesarios, así como la respuesta que nos envía el servidor.

3. TECNOLOGÍAS Y HERRAMIENTAS UTILIZADAS

En este capítulo describiré las tecnologías empleadas durante el transcurso del proyecto y todas las herramientas que he utilizado para el uso de estas tecnologías. Para hacerlo, lo he estructurado en dos puntos: un primer apartado dedicado a las tecnologías y un segundo apartado dedicado a las herramientas utilizadas.

3.1. Tecnologías

3.1.1. PostgreSQL

El sistema de gestión de base de datos que he empleado se llama PostgreSQL, es del tipo objeto-relacional. Es un sistema de gestión de base de datos muy potente y de código abierto, es decir, que el código fuente se encuentra disponible para cualquier usuario. PostgreSQL se encuentra bajo la licencia de los software libres BSD (Berkeley Software Distribution).



Ilustración 5: Cabecera de la página web de PostgreSQL

Entre las ventajas de emplear este sistema de base de datos son las siguientes [2]:

- **Estabilidad y confiabilidad:** las compañías que emplean PostgreSQL nunca han mencionado que este sistema de base de datos haya presentado caídas durante años de operaciones y alta actividad.
- **Extensible:** el código se encuentra disponible para todos sin costes adicionales, por tanto los desarrolladores pueden modificar y personalizar PostgreSQL. Esta es una buena manera de ampliar las funcionalidades del lenguaje.
- **Multiplataforma:** se encuentra disponible en casi cualquier sistema Unix y también cuenta con una versión para Windows.
- **Alto volumen:** la estrategia que adquiere PostgreSQL frente al almacenamiento es el de filas denominado MVCC (Multiversion

Concurrency Control). Esta estrategia consigue una mejor respuesta frente a grandes volúmenes de datos.

Cabe destacar que existen algunas desventajas en el empleo de este sistema de gestión de base de datos:

- Comparado con otro de los grandes sistemas de gestión de base de datos, MySQL, es mucho más lento en operaciones de inserción y actualización.
- No tiene soporte en línea, hay foros oficiales en los que te responden a dudas, pero no hay ayuda obligatoria.
- Consume más recursos que otros sistemas de gestión de base de datos.
- La sintaxis de sus comandos o sentencias, en algunos casos, no es demasiado intuitiva.

He decidido emplear este sistema de gestión de base de datos, ya que, como he comentado anteriormente, es de tipo objeto-relacional y, por ello, cubre con todas las necesidades del proyecto. Además, este tipo de gestor es el que se utilizó en el segundo curso de la docencia de grado, aunque con el sistema de gestión de base de datos Oracle Database. A continuación, describiré el otro motivo por el que me he decantado por PostgreSQL, su extensión PostGis.

3.1.2. PostGIS

Es una extensión del sistema de gestión de base de datos PostgreSQL, que permite añadir la capacidad de utilizarlo como una base de datos espacial, es decir, añadir las características propias de un Sistema de Información Geográfica (SIG).



Ilustración 6: Icono de la extensión PostGis

Esta extensión ha sido desarrollada por Refrations Research Inc, y está disponible bajo una licencia GLP (General Public License).

Como PostGis ha sido creado a partir del código de PostgreSQL, se beneficia de todas sus características, y además agrega otro tipo de datos, entre otros, los de tipo geométrico, tipo geográfico y mapa de bits. También, añade funciones y operadores para poder aplicarlos a estas nuevas geometrías. Todas estas nuevas características hacen que el sistema de gestión de base de datos espacial sea rápido y robusto.

Para poder utilizar los tipos de datos que he mencionado anteriormente es necesario agregarles un parámetro nuevo denominado SRID (Spatial Reference System Identifier), que es un identificador estándar único que hace referencia a un Sistema de Coordenadas concreto. La extensión PostGis emplea el estándar EPSG (European Petroleum Survey Group) como estándar de identificador para los sistemas de coordenadas, por lo que el identificador SRID es el mismo que se emplea en el estándar EPSG.

Durante el transcurso de los análisis he utilizado dos sistemas de referencia distintos, que son los siguientes:

- EPSG: 4326, es el sistema de referencia espacial WGS84. Este sistema es el utilizado en la primera fase tanto para las fotografías/videos como para las ciudades y sus principales puntos de interés. Es un sistema de coordenadas geográficas (latitud y longitud) mundial, que permite localizar un punto en cualquier lugar de la tierra sin necesidad de tener otro punto de referencia. Las características de este sistema se encuentran en la página web: <http://spatialreference.org/ref/epsg/4326/>.
- EPSG: 25830, es el sistema de referencia ETRS89, en su proyección UTM del huso 30 y zona norte. Es un sistema de referencia cartesiano y sus unidades están expresadas en metros. Se emplea en la segunda fase que consiste en la construcción de buffers en los puntos de interés y el dibujo de los mapas. Sus características se encuentra en la página web: <http://spatialreference.org/ref/epsg/25830/>.

3.1.3. Python

Es un lenguaje de programación, creado por el europeo Guido Van Rossum. Su característico nombre se inspiró en el grupo de cómicos ingleses “Monty Python”, del cual, su creador, era fan.



Ilustración 7: Icono del lenguaje de programación Python

El lenguaje de programación Python está basado en la simplicidad, en la versatilidad y en la rapidez para el desarrollo de sus códigos. Por este tipo de atributos, este lenguaje, ha conseguido una gran expansión y popularidad entre los programadores de todo el mundo.

Las características principales por las que se rige este lenguaje de programación son las siguientes:

- **Propósito general:** es un lenguaje que permite la creación de todo tipo de programas, puede realizar aplicaciones para sistemas operativos, servidores de red, páginas web...
- **Multiplataforma:** aunque en principio se desarrolló para Unix, pero se puede emplear en cualquier sistema operativo compatible, siempre que se haya le instalado un intérprete.
- **Interpretado:** se ejecuta línea por línea siendo traducido por el intérprete, es decir, que el código escrito no es compilado antes de su ejecución. Esto ofrece mayor rapidez en el desarrollo del programa pero mayor lentitud mientras se ejecuta.
- **Orientado a objetos:** Python soporta la programación orientada a objetos, aunque no se considera su fuerte, de esta manera puede crear programas más complejos.
- **Sintaxis clara:** utiliza la indentación (márgenes) como método para separar porciones de código frente a otros lenguajes que emplean llaves o palabras clave (begin y end). De esta manera se consigue un aspecto similar entre los códigos de todos los usuarios.
- **Funciones y librerías:** una de las mejores características es que dispone de una gran cantidad de funciones del propio lenguaje, y también existen muchas librerías creadas y difundidas por los propios usuarios de forma gratuita.

En la actualidad existen dos versiones de Python. La primera es Python 2.x, cuya última versión es la 2.7.12; la segunda es la Python 3.x, cuya última versión es la 3.5.2 [3]. Estas versiones tienen elementos distintos en su forma de escribir el código, por lo que una aplicación producida en una versión no funciona si se ejecuta en la otra. Para evitar este inconveniente, existen herramientas que convierten el código de una a la otra.

Finalmente, he optado por este lenguaje, ya que se trata de un código mucho más sencillo y limpio que el de sus homólogos (Java, PHP, C++...), y su curva de aprendizaje es mucho más rápida. Concretamente, he optado por emplear el lenguaje es su versión 2.7, ya que la considero mucho más asentada que su hermana mayor 3.5.

3.2. Herramientas

3.2.1. ArcGis

Es un conjunto de componentes (ArcCatalog, ArcMap, ArcToolbox, ArcGlobe) que se encuentra dentro del ámbito de los SIG. La plataforma ha sido creada, producida y distribuida por la empresa Esri (Environmental Systems Research Institute), fundada en 1969.

Este sistema que permite al usuario recopilar, organizar, administrar, analizar, compartir y distribuir información geográfica. Por lo que es considerada la plataforma líder en su campo.

Existen distintos tipos de ArcGis: el primero, ArcGis Desktop, para los equipos de escritorio y portátiles; el segundo, ArcGis for Server, para utilizar sus herramientas mediante navegadores web; y en tercer lugar, ArcGis Mobile, perfecto para dispositivos móviles como smartphones y tabletas.



Ilustración 8: Tipos de plataformas ArcGis

3.2.2. PgAdmin III

Es una herramienta que permite administrar bases de datos relacionales en PostgreSQL y una de las aplicaciones más completa dentro de las herramientas de código abierto.

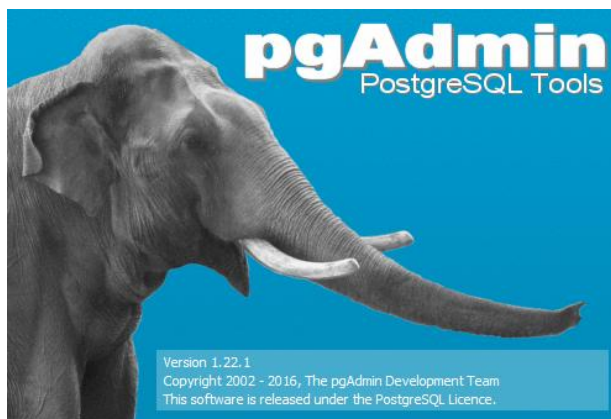


Ilustración 9: Pantalla de carga de PgAdmin III

Cuenta con una ventana gráfica que facilita enormemente la administración de las bases de datos. La aplicación incorpora un editor SQL (Structured Query Language), con el que se pueden ejecutar las consultas automáticamente. Está diseñado para cubrir las necesidades de todos los usuarios, desde hacer consultas simples hasta gestionar bases de datos realmente complejas.

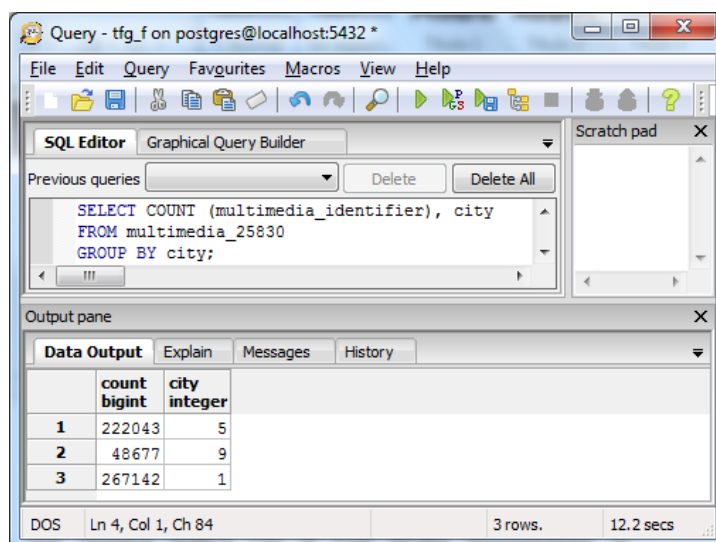


Ilustración 10: Editor SQL con consulta de ejemplo

3.2.3. IDLE (Python GUI)

IDLE (Integrated DeveLopment Environment) es el entorno de desarrollo que se obtiene al descargar Python en su versión 2.7 desde su página web oficial (<https://www.python.org/>). Es una herramienta completamente programada con el lenguaje Python, para hacerlo se utilizó la librería Tkinter, que permite la creación de ventanas gráficas. Con IDLE se pueden crear y ejecutar todo tipo de programas de forma rápida, fácil e intuitiva. Entre sus características destaco las siguientes:

- Editor de texto multiventana.
- Destacado de la sintaxis propia del lenguaje.
- Función de autocompletado.
- Depurador con posibilidad de paso a paso y puntos de interrupción.

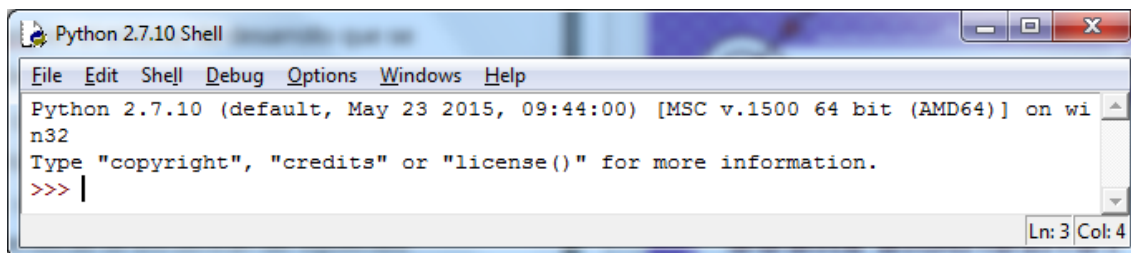


Ilustración 11: Ventana principal IDLE

3.2.4. Dia

Es un programa de software libre, cuyo propósito general es la creación de diagramas. Consta de diferentes módulos, cubre diferentes tipos de diagramas. Entre estos módulos se encuentran, la creación de diagramas de entidad-relación, diagramas UML, diagramas de flujo, diagramas de flujo, diagramas de circuitos eléctricos, etc.

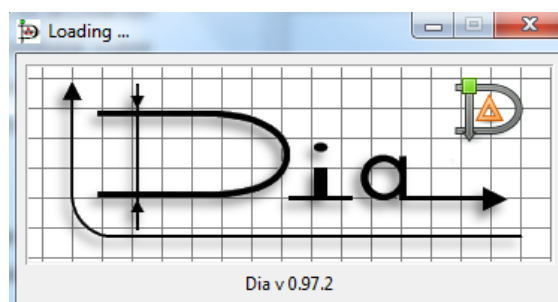


Ilustración 12: Dia. Editor de diagramas

He utilizado este programa por su sencillez y por ser altamente intuitivo en la creación de diagramas. Aunque he utilizado el programa en su forma básica, tiene un gran potencial conseguido gracias a sus herramientas.

Con esta aplicación he implementado el diseño conceptual de la base de datos y el esquema de la aplicación desarrollada.

4. BASE DE DATOS

He dividido este capítulo en dos apartados claramente diferenciados. El primero, lo he destinado a explicar de dónde he obtenido todos los datos que emplearé en el análisis y la obtención de una conclusión. En el segundo apartado, explicaré los conceptos más relevantes de la base de datos y sus entidades.

4.1. Fuente de los datos y datos de partida

Este primer apartado está dividido en otras tres partes, la primera detalla la recogida de la información de las fotografías de Flickr; la segunda, explica cómo he conseguido la cartografía de cada una de las ciudades; y la tercera, muestra cómo he obtenido la cartografía de los monumentos.

4.1.1. Datos de las fotografías

En primer lugar, debo mencionar que toda la información referida a las imágenes y videos de la base de datos ha sido proporcionada por el Departamento de Sistemas Informáticos y Computación de la Universidad Politécnica de Valencia (UPV).

El conjunto total de datos es una lista, que posee más de 100 millones de registros repartidos entre fotografías y videos. La lista se ha podido confeccionar gracias a datos abiertos disponibles por el sitio web Yahoo! Flickr. Además, he de decir, que todos estos datos están sujetos a una licencia de Creative Commons [4], es decir, que los datos se pueden emplear en trabajos y estudios siempre y cuando se muestre la fuente de los datos.

Este conjunto de datos se encuentra almacenado en la plataforma Amazon Web Services (AWS) [5] y por tanto la descarga ha de realizarse por esta plataforma. AWS es un servicio de almacenamiento masivo de información que tiene unos costes económicos. Para poder utilizarlo, lo primero que se debe hacer es obtener un usuario, después, hay que que realizar una petición para la descarga, una vez descargados los datos (que tienen como nombre “I3 - Yahoo Flickr Creative Commons 100M (14G) (Hosted on AWS)”), hay que realizar las siguientes operaciones:

1. Descargar e instalar s3cmd desde la página web ‘<http://s3tools.org/download>’. El s3cmd es una herramienta que permite realizar operaciones sencillas desde la consola de comandos. También se pueden emplear otros gestores de descargas pero recomiendo utilizar este por su sencillez y por tener una buena sinergia con AWS.

2. Configurar el gestor de descarga, ejecutando el comando 's3cmd –configure' en la consola de comandos. Una vez ejecutado, lo primero que va a pedir es el "Access Key" y el "Secret Key" de la cuenta de Amazon (para obtenerlos hay que ir a la página web de Amazon²), a su vez, se pueden especificar otras opciones, como la encriptación durante la transferencia, configuración de proxy, etc, opciones que no se van a emplear en este estudio.
3. Ejecutar el comando 's3cmd ls s3://yahoo-webscope/l3set10/' en la consola, con este comando se pueden ver todos los elementos que contiene el conjunto de datos.
4. Ejecutar en la consola de comandos la siguiente instrucción 's3cmd get --recursive s3://yahoo-webscope/l3set10/', con ella se descargan todos los ficheros contenidos.

Con todos los comandos ejecutados en siguiendo este orden, conseguiremos un total de 10 ficheros comprimidos con un formato .bzip2 que se encontraran descargados en el 'dataset'. Para poder ver su contenido hay que descomprimirlos, obteniendo una serie de ficheros con un total de 100 millones de líneas, cada una de estas líneas corresponde a una fotografía o video diferente. Y contiene la siguiente información:

```
* Photo/video identifier
* User NSID
* User nickname
* Date taken
* Date uploaded
* Capture device
* Title
* Description
* User tags (comma-separated)
* Machine tags (comma-separated)
* Longitude
* Latitude
* Accuracy
* Photo/video page URL
* Photo/video download URL
```

² <https://aws-portal.amazon.com/gp/aws/developer/account/index.html?action=access-key>

- * License name
- * License URL
- * Photo/video server identifier
- * Photo/video farm identifier
- * Photo/video secret
- * Photo/video secret original
- * Extension of the original photo
- * Photos/video marker (0 = photo, 1 = video)

Toda esta información, se encuentra en una misma línea, separada por tabulaciones. Cabe mencionar que es posible que alguno de los campos de datos de carezca de información y por lo tanto, hay que considerarlo en el momento de su procesado.

Para poder gestionar correctamente la información descargada, he creado una base de datos en PostgreSQL, con una tabla donde se almacenan todas las características de cada fotografía y video. La tabla consta de muchos campos, por lo que no es posible representarla íntegramente, por lo que solo muestro algunas de sus columnas:

	multimedia_id bigint	user_nsid character varying(51)	user_nickname character varying(150)	date_taken timestamp without time zone	date_uploaded timestamp without time zone	capture_device character varying(1000)
1	801022466	7739465@N04	donrevilo	2007-07-13 18:08:56	2007-07-13 20:00:21	Canon+EOS+400D+DIGITAL
2	4527699946	24429508@N00	Ricardo+Ricote+Rodríguez	2009-08-14 21:05:03	2010-04-17 08:17:25	NIKON+COOLPIX+P6000
3	180538502	49777470@N00	Kwin+y+Lo	2006-06-01 14:19:43	2006-07-03 11:10:24	Hewlett-Packard+hp+PhotoSmart
4	6666405629	56183392@N00	El+Fabianaco	2012-01-02 14:50:04	2012-01-09 13:12:55	Apple+iPhone+4
5	287577914	90356931@N00	alrom	2008-09-20 23:38:42	2008-09-21 16:59:59	Canon+EOS+400D+DIGITAL
6	6875284790	34186246@N08	ddie	2010-11-05 19:22:44	2012-03-27 19:24:48	Apple+iPhone+3GS
7	12396634013	85124698@N00	benjami	2014-02-09 02:02:26	2014-02-09 02:02:26	
8	2272277043	74286186@N00	mitko denev	2008-02-13 19:26:31	2008-02-18 01:19:42	Panasonic+DMC-FX33
9	8333845139	91809844@N08	JOHNJGM	2013-01-01 04:28:38	2013-01-01 20:37:29	SONY+NEX-5
10	5775518588	61346887@N05	Ricard+Aparicio++Bell	2011-05-27 21:36:22	2011-05-30 11:20:59	NIKON+CORPORATION+NIKON+D3
11	5586072212	56540909@N08	ptkvet	2011-03-11 13:59:08	2011-04-03 20:06:30	NIKON+CORPORATION+NIKON+D7
12	1295781705	37581231@N00	fulviet	2007-08-26 15:32:10	2007-09-01 20:44:40	Canon+DIGITAL+IXUS+IIs
13	3447442587	22103855@N03	Famille+Chollier	2005-10-16 19:58:13	2009-04-16 20:11:26	FUJIFILM+FinePix+S602+ZOOM
14	4645441817	42083039@N00	Thundershead	2010-05-24 19:03:41	2010-05-27 23:54:42	NIKON+CORPORATION+NIKON+D3
15	5018015375	11217496@N03	acidka	2010-09-23 23:55:49	2010-09-23 21:55:49	
16	2346095652	60524783@N00	YasT	2008-03-18 21:02:40	2008-03-19 19:38:14	Canon+EOS+400D+DIGITAL
17	5253243927	83746775@N00	Arguez	2010-07-03 19:12:17	2010-12-12 10:02:27	Canon+EOS+40D
18	319736539	21427355@N00	jcrr	2006-12-11 13:52:44	2006-12-11 19:01:27	FUJIFILM+FinePix+Z1
19	7244379804	38193766@N08	Antonio+Tajuelo	2012-05-13 12:53:05	2012-05-21 22:46:59	Canon+PowerShot+Pro1

Ilustración 13: Tabla multimedia con todos sus campos

He creado un código utilizando la aplicación Python para que lea todas las líneas de los ficheros descargados, analice la información contenida en cada campo y la introduzca en la tabla de la base de datos mencionada anteriormente. Los campos que en los ficheros no tienen valor se introducirán en la base de datos con el valor 'null'. A la vez que se añade la información del 'dataset' en la tabla, hay que añadir una nueva información que permite georeferenciar la fotografía o video. Esta nueva información es

la unión de dos de sus campos, la longitud y la latitud, y un tercero que es el código EPSG del sistema de referencia WGS84 (4326). Para poder introducir esta información en el campo que posee características geométricas, hay que utilizar la siguiente función de Postgis:

```
"ST_GeomFromText('POINT('+longitud+" "+latitude+)',4326)";
```

De los 100 millones de fotografías y videos que descargué, he conseguido procesar correctamente un total de 99.982.825. La causa de que no pudiera procesar los demás archivos es que tenían alguno de sus atributos demasiado largo, lo que causaba que el programa no pudiera cortar la información de la forma programada para un archivo estándar.

4.1.2. Cartografía de las ciudades

Era necesario hacer una selección de ciudades y opté por poner algunas ciudades de otros países como Berlín, Lisboa, Londres, París o Roma, ya que se encuentran relativamente cerca de España y son conocidas por la mayoría de los usuarios. También introduje las ciudades más importantes y con mayor número de turistas de España, Madrid, Barcelona, Valencia, Sevilla y Palma de Mallorca.

Una vez seleccionadas las candidatas a ser analizadas surgió el problema de dónde podía obtener la cartografía de los límites del área metropolitana. La opción que he propuesto es la cartografía abierta de Open Street Maps. La descarga de esta página web es muy sencilla y rápida por lo que considero que es la mejor opción para este estudio. Los pasos a seguir para la descarga son los siguientes:

- En primer lugar, hay que dirigirse a la página web de Open Street Maps (OSM³). Para trabajar con esta web no es necesario estar registrado, pero puede ser interesante tener una cuenta para poder disfrutar de todo su contenido.
- Una vez nos encontramos en la página inicial, basta con introducir el nombre de la ciudad cuya geometría queremos descargar y seleccionar la capa de datos que nos interesa. En el visualizador se puede ver la geometría seleccionada y su identificador.

³ <https://www.openstreetmap.org>

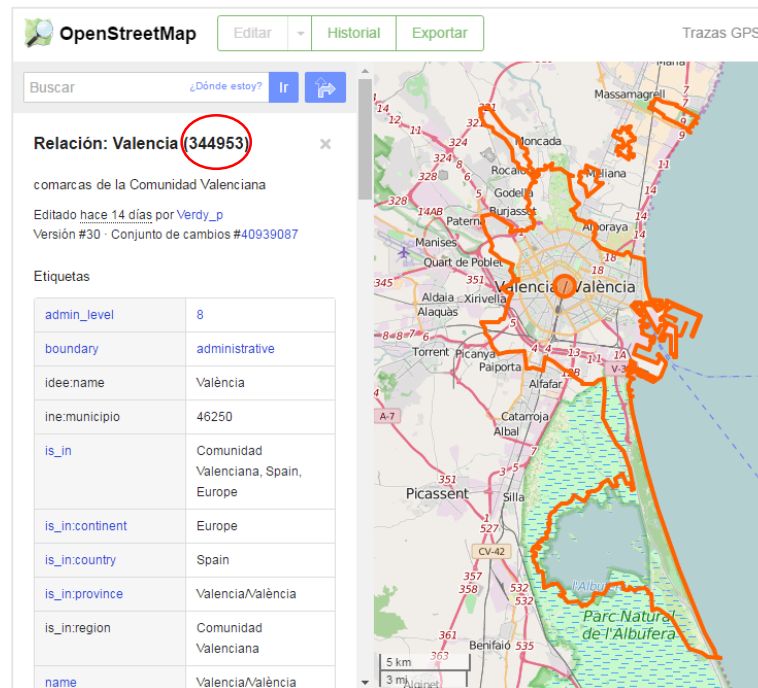


Ilustración 14: Selección de una geometría OSM

- A continuación, hay que acceder a la base de datos de Open Street Maps⁴ e introducir el número identificador de la geometría seleccionada.

Polygon creation

Database was last updated on: 2016-08-04T13:29:02Z

This will generate the whole geometry of the given OSM relation id, with the corresponding sub-relations. When the geometry is available, it is possible to generate simplified geometries from this one, and export them as .poly, GeoJSON, WKT or image formats.

Id of relation

Ilustración 15: Búsqueda de una geometría OSM

- Por último, hay que seleccionar el formato que más se ajuste a las necesidades del trabajo, en este caso, utilizo WKT, un formato de codificación que se emplea para describir objetos espaciales expresados en forma vectorial [6].

⁴ <http://polygons.openstreetmap.fr>

List of available polygons for id = 344953

params ▾	timestamp	NPoints	Length	WKT	GeoJSON	poly	Image
0	2016-05-16 18:38:22.558293	2904	0.324743251996	WKT	GeoJSON	poly	image

Ilustración 16: Descarga de una geometría OSM

Una vez toda la cartografía está descargada y almacenada en ficheros, se crea la tabla que contendrá todos los datos de las ciudades, y cuyo aspecto es el siguiente:

	id integer	name character varying(30)	geom geometry(MultiPolygon,4326)
1	1	barcelona	0106000020E610000006000000010300000001000000067000000990CC7F3197000401E
2	2	berlin	0106000020E61000000100000001030000000100000007F0C0000355C89E53B2D2A4037
3	3	lisboa	0106000020E6100000010000000103000000010000000A4000000FC8F4C874E6B22C0EE
4	4	londres	0106000020E61000000200000001030000000100000006C270000B31E6393A14FE0BF7C
5	5	madrid	0106000020E6100000010000000103000000010000000B50500004B02D4D4B2150FC084
6	6	palma de mallorca	0106000020E610000003000000010300000001000000072070000EEA6A503A38C04406E
7	7	paris	0106000020E61000000100000001030000000100000004A0300000D658C2DA18CB01409F
8	8	roma	0106000020E6100000030000000103000000010000000661000006893C3279D782840D0
9	9	valencia	0106000020E610000004000000010300000001000000077090000C59A801596AEDBBFE7
10	10	sevilla	0106000020E610000002000000010300000001000000026000000CDE7DCED7A2118C087

Ilustración 17: Tabla ciudades en PgAdmin III

4.1.3. Cartografía de los puntos de interés

Para obtener la cartografía de los puntos de interés, lo primero, es realizar una búsqueda de cuáles son los puntos de interés de la ciudad que se va a analizar. En mi caso, he obtenido la información del Portal de Transparencia y Datos Abiertos del Ayuntamiento de Valencia⁵. Estos datos se pueden utilizar con total libertad, sin necesidad de referenciarlos en el documento, ya que han sido elaborados de manera abierta para todo el público.



Ilustración 18: Web del Portal de Transparencia y Datos Abiertos

⁵ <http://gobiernoabierto.valencia.es/es/>

Los datos que ofrece el Ayuntamiento de Valencia se dividen en seis secciones: medio ambiente, sociedad y bienestar, transporte, urbanismo e infraestructuras, salud y turismo.



Ilustración 19: Catálogo de datos abiertos disponibles

Para este trabajo, el catálogo que más interesante resulta es el de turismo, que posee un total de 10 conjuntos de datos, entre los que encontramos la ubicación de: cortes de tráfico por las fallas; carriles bici de la ciudad; barracas de fallas; monumentos falleros; puntos wifi públicos; lavapiés de las playas; pasarelas de las playas; fuentes de agua pública; estaciones de Valenbisi; y monumentos turísticos más importantes.

Para obtener la información requerida se emplea el conjunto de datos de los monumentos turísticos más importantes. La última fecha de actualización es del 25 de Mayo del 2015, por lo que he supuesto que todos los monumentos de la ciudad se encuentran integrados en el conjunto de datos. Las condiciones de uso de estos datos se encuentran bajo la licencia Atribución 4.0 Internacional (CC BY 4.0), que nos permite emplearlos para cualquier fin [7].



Ilustración 20: Conjunto de datos de Monumentos turísticos

Los monumentos y puntos de interés de la ciudad descargados de este portal poseen los siguientes atributos:

- X: coordenada x del monumento turístico.
- Y: coordenada y del monumento turístico.
- Nombre: nombre del monumento turístico.
- Numpol: número de portal.
- Idnotes: identificador único de infocuidad (equipamientos municipales)
- Codvia: código de la vía donde se encuentra.
- Teléfono: número de teléfono.
- Ruta: ruta de agrupación, ruta romana, medieval, contemporánea, marinera...

Para el propósito del análisis, la mayoría de los atributos que nos ofrece el portal de datos abiertos no es relevante, ya que únicamente interesa conocer la posición del monumento para poder georreferenciarlo. Por eso, los campos que he empleado, son las coordenadas centrales del monumento (atributo X e Y) y el nombre del monumento.

La descarga de la información se puede realizar en diversos formatos: wfs, csv, geojson, shape, gml, wms, kml o kmz. Para este proyecto he optado por descargarlo en formato csv ya que posteriormente he de tratar la información.

En el fichero de descarga hay un total de 225 registros, por lo que he realizado un filtrado, ya que, hay lugares que no son significativos para el análisis. Tras el filtrado, el fichero queda con un total de 86 puntos de interés.

Una vez realizado este proceso, cada uno de los monumentos queda representado mediante un punto (x e y). Esto representa un problema, ya que a partir de los puntos que disponemos debo encontrar el área que forma cada uno de los monumentos, que suele ser un elemento poligonal. Para solucionar el problema el departamento de Sistemas Informáticos y Computación ha creado dos programas en lenguaje java. El primer programa contiene la información de los monumentos que he descargado:

```
String[][] coordinates={{"725500.400543468887918","4372705.106547094881535","LONJA"},
{"725459.399893339956179","4372610.107194671407342","MERCADO CENTRAL"},
{"725762.401831573224626","4372925.104410943575203","MONUMENTOS DE LA PLAZA DE LA VIRGEN"},
{"725625.394661780679598","4371868.110010265372694","ESTACION DEL NORTE"},
{"725777.930779632763006","4372849.758384290151298","CATEDRAL"},
{"725690.404039961867966","4373241.103233434259892","TORRES DE SERRANOS"},
{"727734.39187064988073","4370759.101072469726205","CIUDAD ARTES Y CIENCIAS"},
{"725426.405622537597083","4373407.103372258134186","MUSEO BENLLIURE"},
{"725261.396756500587799","4372175.11033939756453","MUVIM"};}
```

Ilustración 21: Lista de monumentos y sus coordenadas

A su vez, realiza una llamada al otro programa para poder obtener la información requerida mediante estas dos líneas:


```
RC = accesoCatastro.readServiceRC(coordinates[i][0], coordinates[i][1]);
parcel = accesoCatastro.readServiceParcel(RC);
```

Ilustración 22: Conexión con el segundo programa

El segundo programa se encarga, primero, de recoger las coordenadas de cada uno de los puntos de interés de la ciudad, y, posteriormente, realizar una conexión con la web de Catastro [8].

```
String addr="http://ovc.catastro.meh.es/ovcservweb/OVCSWLocalizacionRC/OVCCoordenadas.asmx/Consulta_RCCOOR?SRS=EPSG:25830&Coordenada_X=" + X +
"&Coordenada_Y=" + Y;
```

Ilustración 23: URL de Catastro con coordenadas

A partir de esta conexión e introduciendo como argumentos de entrada las coordenadas del monumento, se obtiene su referencia catastral.

Una vez tenemos la referencia catastral de cada elemento, hay que volver a realizar una conexión con el servidor de Catastro, pero esta vez se introduce como argumento de entrada la referencia catastral obtenida anteriormente:

```
String addr="http://ovc.catastro.meh.es/INSPIRE/wfsBU.aspx?service=wfs&version=
2&request=getfeature&STOREDQUERY_ID=GETBUILDINGBYPARCEL&37824724104173&refcat=" + RC;
```

Ilustración 24: URL de Catastro con la referencia catastral

Una vez se conecta de nuevo con el servidor de Catastro utilizando la referencia catastral, se puede obtener toda la información de la parcela catastral, es decir, las coordenadas de cada uno de los vértices que compone el polígono de la parcela. De esta manera, ya se puede introducir la información de cada uno de los monumentos en una nueva tabla en la base de datos, la información introducida tiene el siguiente aspecto:

	poi_id integer	poi_nombre character varying(100)	poi_tipo_lugar character varying(20)	poi_poligono geometry(MultiPolygon,25830)
1	87	ANTIGUO HOSPITAL GENERAL	monument	0106000020E664000001000000010300000001
2	88	L'ALMOINA. CENTRO ARQUEOLÓGICO	monument	0106000020E664000001000000010300000001
3	89	ATARAZANAS	monument	0106000020E664000001000000010300000001
4	90	MUSEO BENLLIURE	museum	0106000020E664000001000000010300000001
5	91	ERMITA DE SANTA LUCÍA	monument	0106000020E664000001000000010300000001
6	92	EDIFICIO DEL RELOJ (FUERTO)	monument	0106000020E664000001000000010300000001
7	93	REAL COLEGIO DE CORPUS CHRISTI (PATRIARCA)	monument	0106000020E664000001000000010300000001
8	94	PALACIO ARZOBISPAL	monument	0106000020E664000001000000010300000001

Ilustración 25: Tabla de monumentos en PgAdmin III

4.2. Diseño de la base de datos

Después de obtener todos los datos necesarios para poder llevar a cabo el análisis turístico, he realizado el diseño de la base de datos. En este apartado, explico cómo he almacenado los datos dentro de las distintas tablas, y cuál será el formato de los atributos de cada elemento, así como las restricciones y características especiales de cada uno de ellos.

Como ya he expuesto anteriormente, he utilizado un modelo de base de datos relacional, concretamente PostgreSQL, para poder gestionar la base de datos, ya que este utiliza el lenguaje SQL (Structured Query Language), que permite el acceso a las bases de datos de tipo relacional y operar de manera sencilla sobre ellas. También he usado la extensión PostGIS, que permite añadir elementos de tipo espacial a la base de datos.

Este apartado se divide en tres fases. La primera es el diseño conceptual que se centra únicamente en la estructura que tiene la información en la base de datos. La segunda, el diseño lógico, con el que establecemos la relación que existe entre los atributos. Por último, el diseño físico, que diseña las tablas de la base de datos.

4.2.1. Diseño conceptual

Para la implementación del diseño conceptual, de los modelos que existen he utilizado el modelo de entidad-relación. El objetivo es la representación de un diagrama entidad-relación de todos los componentes de la base de datos. Lo primero que se tiene que hacer es identificar todas las entidades y las relaciones que existen entre ellas.

Entidades

Se decide qué información será necesaria para el análisis y se realiza una pequeña definición del objeto.

- **Multimedia:** representa el objeto fotografía o video de los datos de Flickr. Es el componente básico de la web de Flickr. Contienen una gran cantidad de información que es tratada más adelante.
- **City:** es un área urbana en la que funciona fundamentalmente la industria y los servicios. Se designa de esta manera al lugar que tiene una determinada entidad político-administrativa.

- **Punto_interes:** es una obra con un valor suficiente para un determinado grupo humano. Esta obra ha de ser pública. Puede comprender cualquier construcción de tipo artístico, arqueológico o histórico.

En el diagrama entidad-relación que presento al final de este capítulo se representa mediante el siguiente dibujo:

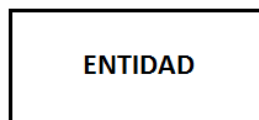


Ilustración 26: Entidad del diseño conceptual

Atributos

Los atributos se refieren a las propiedades que poseen las entidades. Existen dos tipos de atributos, los simples y los compuestos. La diferencia radica en que los simples solo tienen un componente que no se puede dividir, y el compuesto puede tener varios componentes con significados propios. A continuación nombro los atributos de cada entidad:

- **Multimedia:** multimedia_identifier, user_nsid, user_nickname, date_taken, date_uploaded, capture_device, title, description, user_tags, machine_tags, longitude, latitude, accuracy, multimedia_page_url, multimedia_download_url, license_name, license_url, multimedia_server_id, multimedia_farm_id, multimedia_secret, multimedia_secret_original, original_extension, multimedia_marker, geom, city, turist.
- **City:** id, name, geom.
- **Punto_interes:** poi_id, poi_nombre, poi_tipo_lugar, poi_poligono.

En el diagrama entidad-relación represento más adelante, los atributos están simbolizados de la siguiente manera:



Ilustración 27: Atributo del diseño conceptual

Relaciones

En este apartado defino las relaciones que existen entre las entidades nombradas anteriormente, en el caso de la base de datos que he creado solo existe una relación entre dos entidades.

- **Pertenece:** es la relación existente entre la entidad multimedia y la entidad city. En este caso una ciudad puede contener uno o muchos objetos multimedia.

En el diagrama entidad-relación presentado a continuación queda reflejada mediante la siguiente simbología.

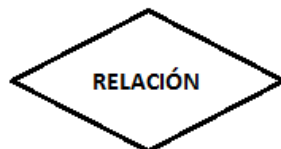


Ilustración 28: Relación del diseño conceptual

Diagrama Entidad-Relación

Es el objetivo final de la primera fase, crear un diagrama entidad-relación con el que se pueda, obtener toda la información de las entidades, sus atributos y relaciones. El esquema que propongo es el siguiente:



Ilustración 29: Diagrama Entidad-Relación del esquema conceptual

4.2.2. Diseño lógico

En esta fase del diseño de la base de datos, parto del resultado del diseño conceptual, y este lo transformo de manera que se adapte a la tecnología que voy a utilizar. La base de datos que quiero implementar, utiliza un sistema de gestión de base de datos (SGBD) relacional, por lo que en esta etapa tengo que obtener el conjunto de relaciones con sus atributos, claves primarias y claves foráneas.

Creación de tablas

En este apartado creo las tablas del esquema lógico, a partir de las entidades del esquema conceptual diseñado anteriormente.

Para definir las tablas, seguiré las siguientes reglas: el nombre de las tablas se escribe en mayúsculas y negrita, los atributos de las tablas se encierran dentro de paréntesis, la clave primaria se encuentra subrayada y en negrita, y por último la clave

ajena se encuentra en cursiva y negrita. Las siglas CP significan clave primaria y CAj es la abreviatura de clave ajena o foránea.

MULTIMEDIA (**multimedia identifier**, user_nsid, user_nickname, date_taken, date_uploaded, capture_device, title, description, user_tags, machine_tags, longitude, latitude, accuracy, multimedia_page_url, multimedia_download_url, license_name, license_url, multimedia_server_id, multimedia_farm_id, multimedia_secret, multimedia_secret_original, original_extension, multimedia_marker, geom, **city**, turist)

CP = { **multimedia identifier** }

CAj = { **city** } → CITY

CITY (**id**, name, geom)

CP = { **id** }

PUNTO_INTERES (**poi_id**, poi_nombre, poi_tipo_lugar, poi_poligono)

CP = { **poi_id** }

En la tabla de multimedia existe una clave ajena, este atributo no puede contener valores nulos, ya que, se trata de la clave principal de otra entidad. Por otra parte, tanto la modificación como el borrado se realizan en cascada, lo que quiere decir que si se modifica o borra, la acción se propaga.

Clave ajena	Nulo	Modificación	Borrado
city --> CITY (id)	No	Cascada	Cascada

Tabla 1: Clave ajena de la entidad multimedia

4.2.3. Diseño físico

En esta fase del diseño de la base de datos, transformo la estructura obtenida en la fase anterior, el objetivo primordial es conseguir una mayor eficacia en el sistema de gestión de base de datos, como ya se ha mencionado anteriormente se emplea PostgreSQL con la extensión PostGIS para otorgarle la capacidad espacial.

Creación de las tablas

En este apartado expongo el contenido de cada una de las tablas de la base de datos y el tipo de dato de cada atributo de la entidad. También muestro la sentencia para la creación de las tablas empleando la sintaxis propia de SQL (Structured Query Language).

TABLA MULTIMEDIA: contiene toda la información referida a las fotografías o videos descargados de la plataforma de Amazon y ofrecida por Flickr.

Atributos de la entidad:

Nombre	Tipo de dato	Tamaño	Nulos	Valor por defecto
multimedia_identifier	BIGINT	-	NO	-
user_nsid	CHARACTER VARYING	51	SI	-
user_nickname	CHARACTER VARYING	150	SI	-
date_taken	TIMESTAMP WITHOUT TIME ZONE	-	SI	-
date_uploaded	TIMESTAMP WITHOUT TIME ZONE	-	SI	-
capture_device	CHARACTER VARYING	1000	SI	-
title	CHARACTER VARYING	3001	SI	-
description	CHARACTER VARYING	10000	SI	-
user_tags	CHARACTER VARYING	10001	SI	-
machine_tags	CHARACTER VARYING	1500	SI	-
longitude	DOUBLE PRECISION	-	SI	-
latitude	DOUBLE PRECISION	-	SI	-
accuracy	INTEGER	-	SI	-
multimedia_page_url	CHARACTER VARYING	253	SI	-
multimedia_download_url	CHARACTER VARYING	254	SI	-
license_name	CHARACTER VARYING	103	SI	-
license_url	CHARACTER VARYING	255	SI	-
multimedia_server_id	INTEGER	-	SI	-
multimedia_farm_id	INTEGER	-	SI	-
multimedia_secret	CHARACTER VARYING	31	SI	-
multimedia_secret_original	CHARACTER VARYING	32	SI	-
original_extension	CHARACTER VARYING	5	SI	-
multimedia_marker	BIT	1	SI	-
geom	GEOMETRY (POINT, 4326)	-	SI	-
city	INTEGER	-	SI	-
turist	BOOLEAN	-	SI	-

Tabla 2: Atributos de la entidad multimedia

Sentencia SQL para la creación de la tabla:

```
CREATE TABLE multimedia (
    multimedia_identifier bigint NOT NULL,
    user_nsid character varying(51),
    user_nickname character varying(150),
    date_taken timestamp without time zone,
    date_uploaded timestamp without time zone,
    capture_device character varying(1000),
    title character varying(3001),
    description character varying(10000),
    user_tags character varying(10001),
    machine_tags character varying(1500),
    longitude double precision,
    latitude double precision,
    accuracy integer,
    multimedia_page_url character varying(253),
    multimedia_download_url character varying(254),
    license_name character varying(103),
    license_url character varying(255),
    multimedia_server_id integer,
    multimedia_farm_id integer,
    multimedia_secret character varying(31),
    multimedia_secret_original character varying(32),
    original_extension character varying(5),
    multimedia_marker bit(1),
    geom geometry(Point,4326),
    city integer
);

ALTER TABLE ONLY multimedia
    ADD CONSTRAINT multimedia_pkey
    PRIMARY KEY (multimedia_identifier);
ALTER TABLE ONLY multimedia
    ADD CONSTRAINT multimedia_fkey_city
    FOREIGN KEY (city) REFERENCES city(id)
    ON UPDATE CASCADE ON DELETE CASCADE;
```

Ilustración 30: Sentencia SQL para la tabla multimedia

TABLA CITY: es la tabla que contiene los datos de la geometría de las ciudades que se obtuvieron para el análisis.

Atributos de la entidad:

Nombre	Tipo de dato	Tamaño	Nulos	Valor por defecto
id	INTEGER	-	NO	SERIAL
name	CHARACTER VARYING	30	SI	-
geom	GEOMETRY (MULTIPOLYGON, 4326)	-	SI	-

Tabla 3: Atributos de la entidad city

Sentencia SQL para la creación de la tabla:


```
CREATE TABLE city (
  id integer NOT NULL,
  name character varying(30),
  geom geometry(MultiPolygon,4326)
);
CREATE SEQUENCE city_id_seq
  START WITH 1
  INCREMENT BY 1
  NO MINVALUE
  NO MAXVALUE
  CACHE 1;
ALTER SEQUENCE city_id_seq OWNED BY city.id;
ALTER TABLE ONLY city ALTER COLUMN id
  SET DEFAULT nextval('city_id_seq'::regclass);
ALTER TABLE ONLY city
  ADD CONSTRAINT city_pkey PRIMARY KEY (id);
```

Ilustración 31: Sentencia SQL para la tabla city

TABLA PUNTO_INTERES : es la tabla en la que se encuentran las características de cada punto de interés y su geometría.

Atributos de la entidad:

Nombre	Tipo de dato	Tamaño	Nulos	Valor por defecto
poi_id	INTEGER	-	NO	SERIAL
poi_nombre	CHARACTER VARYING	100	SI	-
poi_tipo_lugar	CHARACTER VARYING	20	SI	-
poi_poligono	GEOMETRY (MULTIPOLYGON, 4326)	-	SI	-

Tabla 4: Atributos de la entidad punto_interes

Sentencia SQL para la creación de la tabla:

```
CREATE TABLE punto_interes (
  poi_id integer NOT NULL,
  poi_nombre character varying(100),
  poi_tipo_lugar character varying(20),
  poi_poligono geometry(MultiPolygon,4326)
);
CREATE SEQUENCE punto_interes_poi_id_seq
  START WITH 1
  INCREMENT BY 1
  NO MINVALUE
  NO MAXVALUE
  CACHE 1;
ALTER SEQUENCE punto_interes_poi_id_seq OWNED BY punto_interes.poi_id;
ALTER TABLE ONLY punto_interes ALTER COLUMN poi_id
  SET DEFAULT nextval('punto_interes_poi_id_seq'::regclass);
ALTER TABLE ONLY punto_interes
  ADD CONSTRAINT poi_pk PRIMARY KEY (poi_id);
```

Ilustración 32: Sentencia SQL para la tabla punto_interes

4.3. Conclusiones de la base de datos

Para mejorar la eficiencia y que el sistema no se sature al ejecutar las consultas para la realización del análisis, he creado una nueva tabla en la base de datos con la información disponible sobre las fotos y videos de Madrid, Barcelona y Valencia (alrededor de 540.000 fotografías/videos). La sentencia SQL que se encarga de realizar el traspaso de esta información es la siguiente:

```
INSERT INTO multimedia (multimedia_identifier, user_nsid, user_nickname, date_taken, date_uploaded,
capture_device, title, description, user_tags, machine_tags, longitude, latitude, accuracy,
multimedia_page_url, multimedia_download_url, license_name, license_url,
multimedia_server_id, multimedia_farm_id, multimedia_secret, multimedia_secret_original,
original_extension, multimedia_marker, geom, city)
SELECT m."multimediaIdentifier", m."userNSID", m."userNickname", m."dateTaken", m."dateUploaded",
m."captureDevice", m.title, m.description, m."userTags", m."machineTags", m.longitude,
m.latitude, m.accuracy, m."multimediaPageUrl", m."multimediaDownloadUrl",
m."licenseName", m."licenseUrl", m."multimediaServerId", m."multimediaFarmId",
m."multimediaSecret", m."multimediaSecretOriginal", m."originalExtension",
m."multimediaMarker", m.geom, c.id
FROM multimedia_gis m, city c
WHERE st_contains(c.geom, m.geom) AND
c.name IN ('valencia', 'madrid', 'barcelona');
```

Por otra parte, para crear las tablas del punto anterior, las coordenadas tanto de los puntos (fotografías y videos) como de los elementos poligonales (ciudades y puntos de interés), se encuentran en el sistema de referencia WGS84 con código EPSG (European Petroleum Survey Group) 4326, que tiene la característica de ser coordenadas geográficas de carácter mundial. Para las consultas sobre la base de datos se emplea la proyección UTM de la zona 30 en el sistema de referencia ETRS89 con código EPSG 25830, de esta manera se consigue coordenadas cartesianas expresadas en metros y se pueden realizar de forma más eficaz las áreas de influencia a los puntos de interés.

Con esto quiero decir que, se tienen que transformar las coordenadas de un sistema a otro. Después de realizar alguna prueba con las consultas, he llegado a la conclusión de que son consultas que requieren una gran cantidad de recursos, ya que, se tienen que realizar la transformación 'al vuelo' de todos los puntos y polígonos. La duración en la ejecución de la consulta es extremadamente larga en estos casos, mas de 30 minutos.

Por esta razón, he decidido crear dos nuevas tablas, una para las fotografías y videos y una segunda para los puntos de interés. Las características de diseño son las mismas que las citadas en las fases anteriores, salvo que realizo una transformación de

coordenadas en el campo 'geom'. De manera que el atributo queda definido de la siguiente manera para los puntos (fotografías y videos):

```
geom geometry(Point,25830)
```

Ilustración 33: Atributo de tipo punto con EPSG 25830

Para las entidades que tienen una geometría poligonal, el nuevo atributo queda definido de la siguiente manera:

```
poi_poligono geometry(MultiPolygon,25830)
```

Ilustración 34: Atributo de tipo poligonal con EPSG 25830

Para el traspaso de dato de la tabla multimedia con EPSG 4326 a la nueva tabla con EPSG 25830, copio todos los datos de los atributos y le aplico el comando de PostGIS 'ST_Transform' [9] al atributo que contiene la geometría de tipo punto especificando como parámetros de entrada la geometría y el código EPSG de salida.

```
ST_Transform(geometry, 25830)
```

Ilustración 35: Sentencia de transformación de sistema de coordenadas

Para la tabla de punto de interés realizo la misma operación de transformación de sistema de referencia. Al final de estas operaciones obtengo dos nuevas tablas las cuales reciben el nombre de: multimedia_25830 y punto_interes_25830.

5. APLICACIÓN

En este apartado explico la aplicación que he desarrollado para realizar todo el trabajo de análisis, y que sirve como soporte para que cualquier usuario sin conocimiento previo en programación pueda extraer la información de las gráficas. El lenguaje de programación que he utilizado es el lenguaje interpretado Python, en su versión 2.7.10 de 64 bits. Para programar de una manera más sencilla pero sin que sea automática con el auto-completado, he empleado el IDLE, que se descarga automáticamente con Python, es un entorno de programación simple y destinado a principiantes o programadores poco avanzados, aunque con gran potencial.

5.1. Aspecto de la aplicación

En primer lugar muestro el aspecto de la ventana principal. Le he añadido características personalizadas para crear un entorno más profesional y ameno.

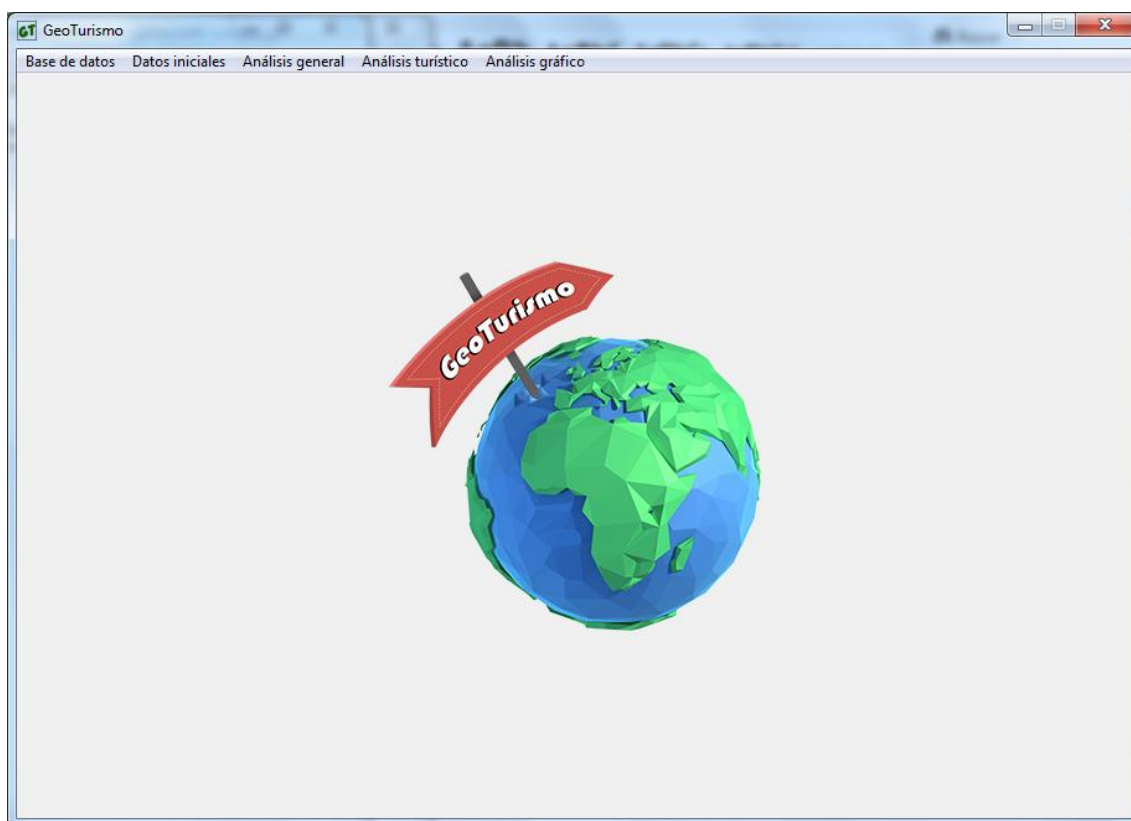


Ilustración 36: Ventana principal de la aplicación

Como se observa, a la ventana gráfica le he incluido una imagen de fondo. Es el logo de la aplicación, y la imagen que caracteriza y le da personalidad a la aplicación.

El logo lo he creado mediante Photoshop, empleando distintas imágenes y adaptándolas entre ellas para obtener una cohesión.

Por otro lado, también le he añadido un icono en la cabecera, que ha sustituido al icono con el que se crea la aplicación automáticamente. También lo he creado a través de la herramienta Photoshop.



Ilustración 37: Icono de la aplicación

5.2. Menús de la aplicación

En este apartado explico los distintos menús desarrollados e implementados en la aplicación.

BASE DE DATOS

La función de este menú es establecer la conexión con la base de datos que he diseñado. Consta de dos opciones, la primera accede a una base de datos predefinida, en este caso el nombre, usuario, contraseña y puerto están ya implementados para poder acceder a mi base de datos. En la segunda opción, es posible acceder a una base de datos cualquiera, únicamente es necesario escribir los datos necesarios para la conexión.

Para poder acceder al resto de menús y obtener resultados, es necesario haber realizado una conexión con alguna base de datos.

DATOS INICIALES

Se encarga de realizar los cálculos más generales, en ellos se analiza toda la base de datos. Este menú sirve de partida para comprobar cuántas imágenes contiene cada ciudad y cuántos usuarios existen en la base de datos.

ANÁLISIS GENERAL

En este menú ya se singularizan los análisis, de manera que se crean gráficas y se obtienen resultados numéricos para Valencia. Existen dos opciones más, la ciudad de Madrid y de Barcelona, pero estas se mantienen desactivadas para evitar que el usuario las pueda pulsar, ya que no están implementadas en el código.

ANÁLISIS TURÍSTICO

Las opciones van dirigidas a realizar análisis con los usuarios que se consideran turistas con los criterios establecidos. Todas ellas muestran un resultado en la ventana gráfica, ya sea a modo de resultado numérico o mediante una gráfica.

ANÁLISIS GRÁFICO

Dentro de este menú, se pueden crear dos series de mapas. La primera, divide en cuatro zonas la ciudad de Valencia, y representa, mediante un mapa de densidad, los lugares en los que más fotografías de usuarios hay por metro cuadrado. La segunda, realiza la misma operación, pero esta vez utiliza solo los usuarios seleccionados como turistas. Los mapas se crean en formato pdf.

5.3. Esquema de la aplicación

A continuación, muestro los directorios de los que consta la aplicación de manera ordenada y gráfica:

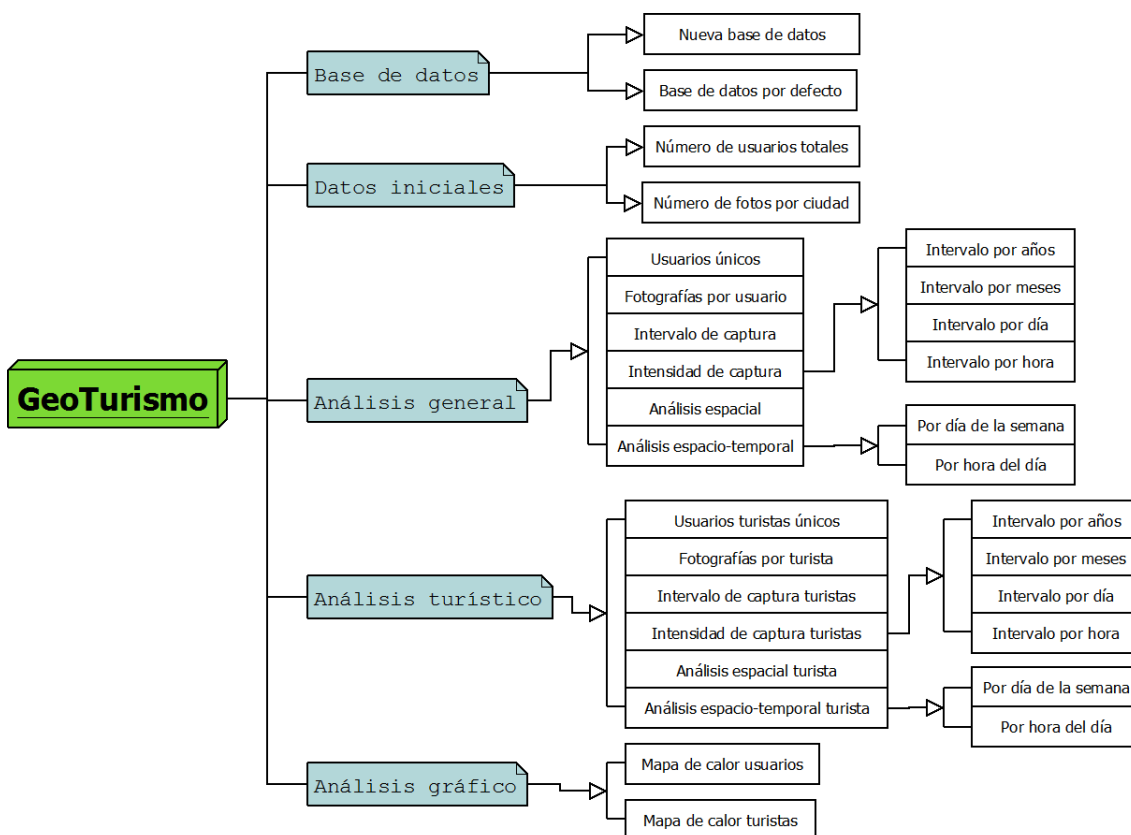


Ilustración 38: Esquema de la aplicación

5.4. Módulos y librerías

En este apartado explico brevemente los módulos y librerías utilizados durante el desarrollo de la aplicación y que son necesarios para realizar todos los cálculos.

En el código fuente quedan definidos de la siguiente manera:

```
from Tkinter import *
import pprint
import psycopg2
import arcpy

import numpy as np
import matplotlib.mlab as mlab
import matplotlib.pyplot as plt
from matplotlib import cm
from matplotlib.backends.backend_tkagg import FigureCanvasTkAgg, NavigationToolbar2TkAgg
from PIL import Image, ImageTk
import time
```

Ilustración 39: Captura de los módulos y librerías utilizadas

Tkinter

Es una librería que viene acompañada con la descarga del propio Python 2.7. La biblioteca otorga la posibilidad de realizar GUI (*graphical user interface* o interfaz gráfica en castellano). Con esta librería he desarrollado la ventana principal, los lienzos en los que inserto tanto el logo como los gráficos, y los distintos menús y botones para la creación de los gráficos [10].

Pprint

Es otra de las librerías que se descargan automáticamente cuando descargas Python. Esta librería permite imprimir por pantalla una gran cantidad de datos y se ha utilizado para lanzar los resultados de las consultas por pantalla. Su nombre viene de la expresión “pretty-print” [11].

Psycopg2

Es un módulo externo descargado, no viene por defecto con Python. Es el adaptador más popular para la base de datos PostgreSQL. Lo he empleado para establecer la conexión con la base de datos y para todas las consultas realizadas una vez conectado [12].



Ilustración 40: Logo Psycopg

Arcpy

Es un módulo muy potente, proporciona una manera de realizar análisis de datos geográficos, convertir datos, administrar datos y automatizar mapas con el lenguaje de programación Python. Para el proyecto, se ha empleado para automatizar, en la medida de lo posible, los mapas de densidad de fotografías y videos [13].

Numpy

Es una biblioteca que agrega mayor soporte para crear vectores y matrices en Python. También posee una gran biblioteca de funciones matemáticas de alto nivel que permite operar con esos vectores y matrices [14].



Ilustración 41: Logo oficial de Numpy

Matplotlib

Es una librería que permite generar gráficos a partir de datos contenidos en listas o arrays en el lenguaje Python. Posiblemente es el módulo más utilizado para realizar gráficos en 2D, ya que, proporciona una manera rápida de visualizar datos y figuras con calidad. Existe una gran cantidad de información en la red de cómo usarlo [15].



Ilustración 42: Logo oficial de Matplotlib

PIL (Python Imaging Library)

Es una librería para el lenguaje de programación Python, añade el soporte necesario para abrir, manipular y guardar una gran cantidad de formatos de archivos de imagen [17]. Esta librería la he utilizado para insertar tanto el logo de la aplicación como el icono.

Time

Es un módulo predeterminado de Python, ofrece la posibilidad de trabajar con fechas y horas. Posee una gran cantidad de formatos para visualizar valores referidos al tiempo [16]. Se ha empleado para obtener automáticamente la fecha de creación de los mapas.

6. ANÁLISIS DE LOS DATOS

En esta parte del trabajo, describo las conclusiones obtenidas durante la realización de todo el proyecto. Para favorecer la interpretación, divido el capítulo en cuatro partes bien diferenciadas. En la primera, realizo un pequeño recuento de las fotografías utilizadas en el análisis, una breve comparación y explicación de porqué utilizo Valencia como ciudad de análisis. La segunda, consiste en un análisis más general, en el cual utilizo todas las fotografías descargadas sin discriminar si un usuario es turista o no. En la tercera, realizo un análisis de la parte de usuarios que considero turistas, los criterios que he seguido para su obtención los expondré más adelante. Por último, realizo una comparativa de los resultados que he obtenido en los dos primeros apartados para obtener una conclusión a partir de ellos.

Por otra parte, para realizar el análisis de los datos he creado gráficas, ya que, una vez realizada la consulta, el resultado suele ser una tabla con varias filas y columnas, las cuales son ininteligibles si no se realizan gráficos. Para un usuario es mucho más sencillo y rápido interpretar un gráfico que una tabla con una gran cantidad de datos. La creación de los gráficos los he realizado mediante un módulo de Python llamado Matplotlib, es relativamente sencillo de utilizar y se pueden conseguir gráficas realmente buenas y llamativas. Para crear los gráficos he establecido una conexión con la base de datos y a partir de ella, ejecutar las consultas en lenguaje SQL para obtener un resultado. Una vez extraída la información de la base de datos, la proceso para poder plasmarla en las gráficas.

También he creado diversos mapas para poder interpretar las consultas espaciales. Los mapas que he creado son mapas de calor. Para la composición de los mapas he utilizado la herramienta de escritorio ArcGis y su módulo de Python llamado Arcpy.

A continuación expongo los diversos análisis realizados con los datos descargados de la red social Flickr.

6.1. Datos iniciales

Son los datos que se pueden extraer de toda la base de datos, consta de tres ciudades distintas. Con los resultados de estos análisis iniciales selecciono la ciudad a la que dedicaré un análisis más exhaustivo.

6.1.1. Número de usuarios totales

El primer análisis es saber cuántos usuarios han publicado alguna fotografía en alguna de las ciudades seleccionadas en la base de datos. Es un dato importante, ya que, a partir de este dato podré calcular otros parámetros de los usuarios.

Lo que realizo en la consulta es un recuento de los identificadores de usuario, de manera que únicamente cuenta los usuarios que no se repiten en la base de datos. La consulta tiene la siguiente forma:

```
SELECT COUNT (DISTINCT user_nsid)
FROM multimedia_25830;
```

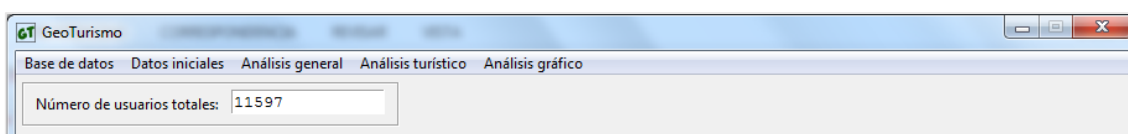


Ilustración 43: Pantalla de número de usuarios totales

El resultado de la consulta es que existen un total de 11597 usuarios distintos en la base de datos.

6.1.2. Número de fotografías por ciudad

Como he comentado anteriormente (apartado 4.3 Conclusiones de la base de datos), son tres las ciudades españolas que he preseleccionado para realizar el análisis turístico. Las ciudades son Madrid, la capital del país, Barcelona por ser una de las ciudades más importantes y con mayor tránsito de turistas y por último Valencia por ser también una ciudad con un gran acogimiento turístico, sobretodo estival [18].

La consulta que he realizado sobre la base de datos para obtener el número de fotografías de cada una de las ciudades es la siguiente:

```
SELECT COUNT (multimedia_identifier), city
FROM multimedia_25830
GROUP BY city;
```

El resultado es una tabla con dos columnas en las que aparecen la cantidad de fotografías agrupadas por ciudad a la que corresponden.

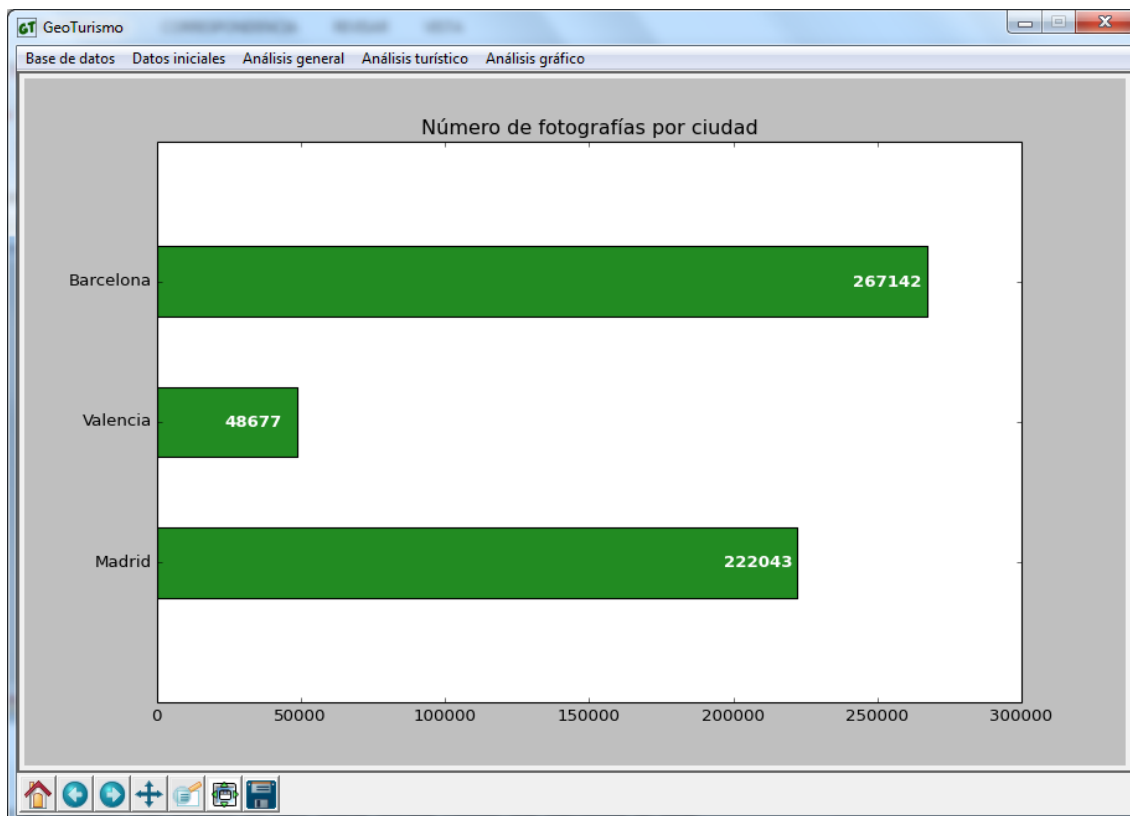


Ilustración 44: Gráfica de número de fotografías por ciudad

Como se aprecia en el gráfico, existe un elevado número de fotografías situadas en las ciudades más grandes del país. Puede resultar curioso que Barcelona se sitúe por delante de Madrid en cantidad de fotos, esto puede deberse a que en Barcelona se encuentran grandes obras de Gaudí, como la Sagrada Familia o la Pedrera, por albergar en plena ciudad la playa de la Barceloneta y por contar con buenas temperaturas durante gran parte del año.

No obstante Madrid no se queda atrás, como refleja la cantidad de fotos de la gráfica. En esta ciudad se pueden visitar lugares emblemáticos como el Palacio Real, el Museo del Prado, el Parque del Retiro, etc.

Por último, se encuentra la ciudad de Valencia que, aunque el número de fotografías se sitúa muy por detrás de las anteriores ciudades mencionadas, cuenta con grandes reclamos turísticos, entre otros: visitar la Ciudad de las Artes y las Ciencias; comer una paella en sus locales típicos o darse un baño en sus grandes playas [19].

Para el análisis he optado por la ciudad de Valencia, que aunque tenga un número menor de fotografías, existen las suficientes como para obtener unas buenas conclusiones, y por otro lado, también se cuenta con un mayor conocimiento de la ciudad y todos sus monumentos y lugares de interés, por lo que resultará más sencillo el tratamiento de los datos.

6.2. Análisis general

Una vez seleccionada la ciudad a analizar, ya estoy en situación de comenzar con el primer análisis, consiste en realizar consultas sobre la base de datos con un carácter más generalista, en un principio se trata de consultas sencillas que se van complicando a medida que se van necesitando resultados más concretos. Como ya dije, las consultas las realizo a la población total de los datos de la ciudad que nos concierne, esto implica que existen datos tanto de usuarios turistas como usuarios locales.

A continuación paso a describir y comentar cada uno de los análisis generales realizados.

6.2.1. Número de usuarios únicos en Valencia

El primer análisis es obtener cuántos usuarios han publicado alguna fotografía en la ciudad de Valencia. Es un dato importante, ya que, a partir de este dato se pueden calcular otros parámetros de los usuarios.

Lo que realizo en la consulta es un recuento de los identificadores de usuario, de manera que únicamente cuenta los usuarios que han publicado en Valencia y no se repiten en la base de datos. La consulta tiene la siguiente forma:

```
SELECT COUNT (DISTINCT user_nsid)
FROM multimedia_25830
WHERE city = 9;
```

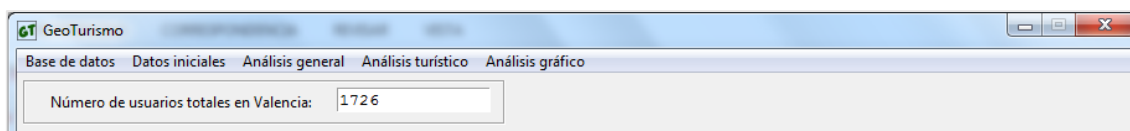


Ilustración 45: Pantalla de número de usuarios en Valencia

El resultado de la consulta es que, para la ciudad de Valencia existen un total de 1726 usuarios distintos que han realizado alguna publicación.

6.2.2. Número de fotografías por usuario

Como he comentado en el primer análisis de este capítulo, la ciudad de Valencia consta de un total de 48677 fotografías publicadas por 1726 usuarios distintos, pero todos los usuarios no capturan el mismo número de fotografías, por lo que es necesario saber la cantidad de imágenes que ha capturado cada usuario y a partir de esto poder averiguar los estadísticos descriptivos de la población.

En primer lugar he obtenido una gráfica de toda la muestra completa para analizar su comportamiento, y a continuación decidir cómo proceder con el análisis. Para ello he realizado la siguiente consulta sobre la base de datos:

```
SELECT COUNT (multimedia_identifier)
FROM multimedia
WHERE city = 9
GROUP BY user_nsid
ORDER BY COUNT (multimedia_identifier) DESC;
```

El resultado lo presento de manera sencilla en una gráfica de Microsoft Excel para llegar a una conclusión de manera rápida y sencilla.

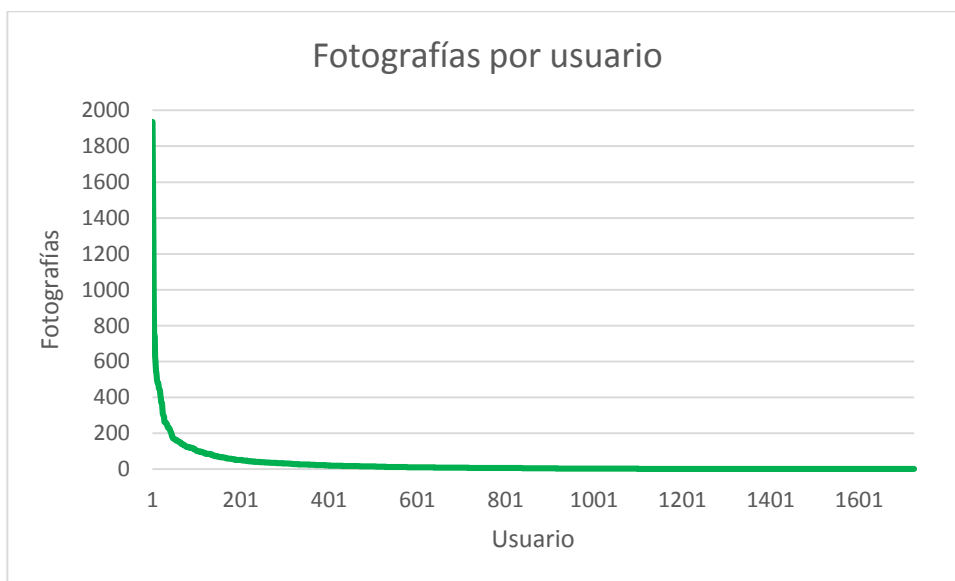


Ilustración 46: Gráfica fotografías por usuario

Los datos estadísticos extraídos de la población completa son los que siguen: la media de fotografías publicadas por usuario es de 6.2, la mediana es de 5.0 y la desviación típica es de 95.5. Se puede apreciar rápidamente que existe una pequeña parte de usuarios que publican una gran cantidad de fotografías, pero el número de publicaciones se reduce drásticamente y la mayoría de los usuarios se sitúan en un número de publicaciones bajas.

Se concluye que para realizar el análisis se tiene que dividir la población total en dos partes, esto lo he tenido que realizar así dado que en la gráfica el número de fotografías publicadas por usuarios disminuye drásticamente y no se puede interpretar correctamente.

La primera consulta corresponde a los usuarios más activos dentro de la aplicación, es decir, los que han publicado en su perfil un número mayor a 50 fotografías

de la ciudad de Valencia. La consulta es algo más compleja que las anteriores y queda definida de la siguiente manera.

```
SELECT COUNT (multimedia_identifier)
FROM multimedia_25830
WHERE city = 9
GROUP BY user_nsid
HAVING COUNT (multimedia_identifier) > 50
ORDER BY COUNT (multimedia_identifier) DESC;
```

La siguiente consulta corresponde con los usuarios menos activos dentro de la plataforma, en este caso he considerado que un usuario es poco activo cuando ha publicado un número igual o inferior a 50 fotografías de una misma ciudad. La consulta es similar a la anterior, únicamente cambio el condicionante.

```
SELECT COUNT (multimedia_identifier)
FROM multimedia_25830
WHERE city = 9
GROUP BY user_nsid
HAVING COUNT (multimedia_identifier) <= 50
ORDER BY COUNT (multimedia_identifier) DESC;
```

El resultado de esta consulta lo transformo en gráfica para ver la tendencia que adquiere la publicación de estos usuarios.

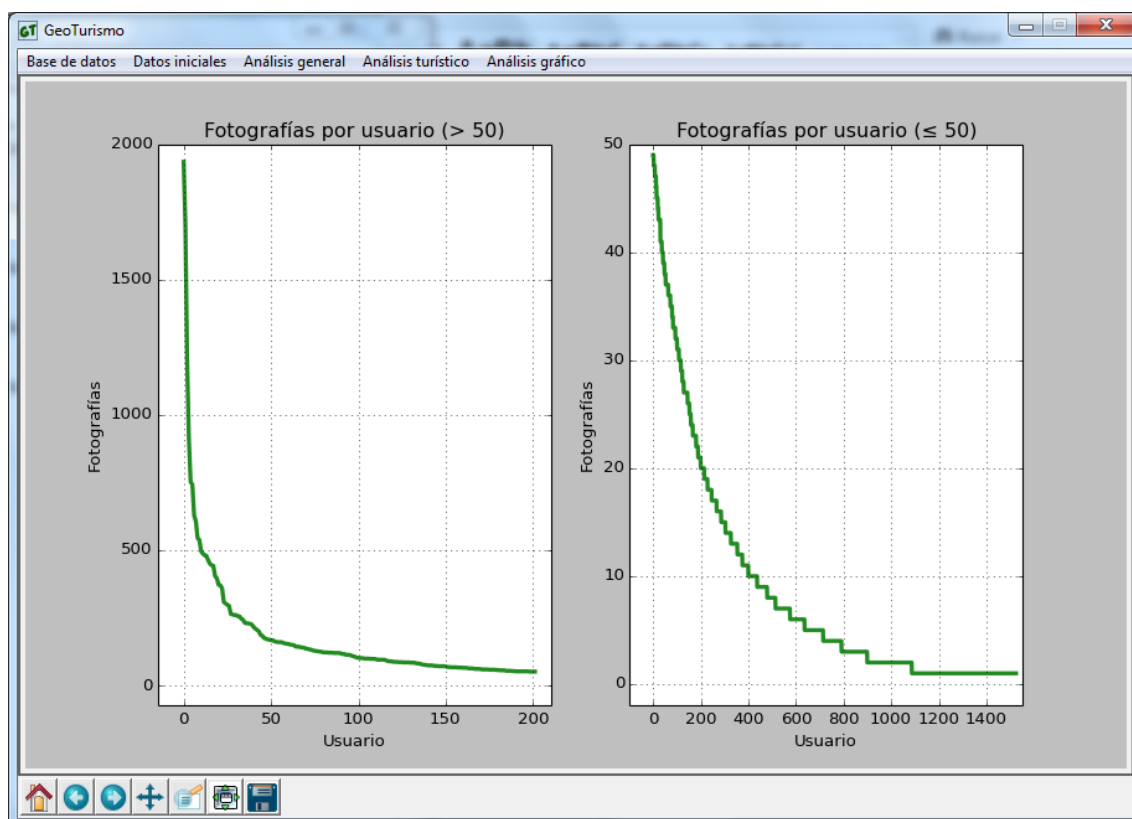


Ilustración 47: Gráfica fotografías por usuarios divididos en dos grupos

Para los usuarios que han publicado un número de fotografías mayor que 50 (gráfica izquierda de la ilustración 47) la media es de 122.7 fotografías, la mediana obtenida es de 101.5 y la desviación típica corresponde a 229.5. Como se puede ver mediante los estadísticos básicos, la muestra seleccionada se corresponde con una poco homogénea y por eso la desviación típica es tan elevada, existen unos pocos usuarios que han publicado un gran número de imágenes y rápidamente disminuyen y la gráfica se estabiliza.

Para los usuarios que han publicado un número de fotografías menor que 50 (gráfica derecha de la ilustración 47) puedo comentar que la media de fotografías publicadas por usuarios es de 4.2 y su mediana es de 4.0, la desviación típica correspondiente a esta muestra es de 10.8. Por otro lado, también se puede observar como la línea tiende rápidamente a 1 fotografía, ya que la gran mayoría de los usuarios publican pocas imágenes.

6.2.3. Intervalo de captura de fotografías

En este análisis pretendo obtener durante cuánto tiempo se han capturado fotografías en la ciudad de Valencia. Para realizar este cálculo he empleado dos métodos distintos.

El primer, y el que empleamos en la aplicación, es la utilización de la función predefinida 'AGE' de PostgreSQL, es capaz de devolver el intervalo de tiempo dado como argumento dos parámetros de entrada [20]. En este caso basta con introducir como parámetros la fecha mínima y la fecha máxima que contienen las fotos de la ciudad de Valencia. La consulta queda formulada de la siguiente manera:

```
SELECT AGE(MAX(date_taken),MIN(date_taken))
FROM multimedia_25830
WHERE city = 9;
```

El resultado de esta consulta, es que desde la primera captura hasta la última hay un lapso de tiempo de 81 años, 10 meses y 1 día. Indagando un poco más en la base de datos se puede ver que hay algunas fechas que pueden ser erróneas, por ejemplo una fotografía capturada en el año 2018, por lo que aplico un filtro. Establezco una fecha mínima en el año 2000 (hay unas pocas fotografías capturadas antes del año 2000) para evitar errores por defecto y una fecha máxima en el año 2016 para eliminar los errores por exceso.

```
SELECT age(MAX(date_taken),MIN(date_taken))
FROM multimedia_25830
WHERE city = 9
AND EXTRACT (year FROM date_taken) > 2000
AND EXTRACT (year FROM date_taken) < 2016;
```

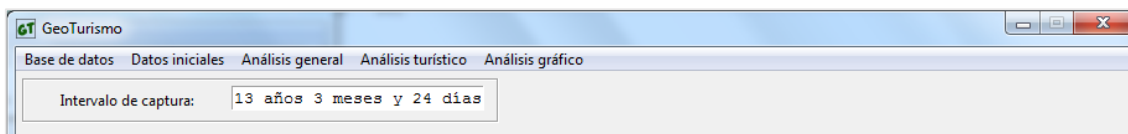



Ilustración 48: Pantalla de intervalo de captura

Con este filtro elimino los posibles errores groseros que puedan desvirtuar la muestra, de esta manera el intervalo de captura entre la primera foto y la última es de 13 años, 3 meses y 24 días.

El segundo método consiste en el empleo de otra función predefinida de PostgreSQL llamada 'EXTRACT', la función devuelve un resultado dependiendo del argumento de entrada [21]. Le introducimos la palabra clave 'day', de esta manera la respuesta obtenida por la función es el intervalo completo en días, también se tiene que introducir el rango de tiempo, que como en el caso anterior, es la primera y última fotografía capturada. La consulta se define de la siguiente manera:

```
SELECT EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken)))
FROM multimedia_25830
WHERE city = 9
      AND EXTRACT (year FROM date_taken) > 2000
      AND EXTRACT (year FROM date_taken) < 2016;
```

El resultado de la consulta es que contamos con un intervalo de captura de imágenes de 4862 días.

6.2.4. Intensidad de captura de fotografías

En esta parte del análisis muestro durante qué periodo de tiempo se han capturado las fotografías de los usuarios y en qué puntos temporales han sido más frecuentes sus capturas. Para realizarlo, divido el análisis en cuatro partes, la primera de ellas es la intensidad de captura separada por años, en la segunda agrupo las imágenes por su mes de toma, la tercera las agrupa por días de la semana y la última por horas del día.

Para realizar esta consulta a la base de datos, utilizo nuevamente el comando 'EXTRACT' para obtener el año en el que se tomó la fotografía. La consulta es la siguiente:

```
SELECT EXTRACT(year FROM date_taken) AS año,
      COUNT(date_taken) AS cantidad_fotos
FROM multimedia_25830
WHERE city = 9
      AND EXTRACT (year FROM date_taken) > 2000
      AND EXTRACT (year FROM date_taken) < 2016
GROUP BY año
ORDER BY año ASC;
```

El resultado de esta consulta es una agrupación de las fotografías de todos los usuarios por su año de captura.

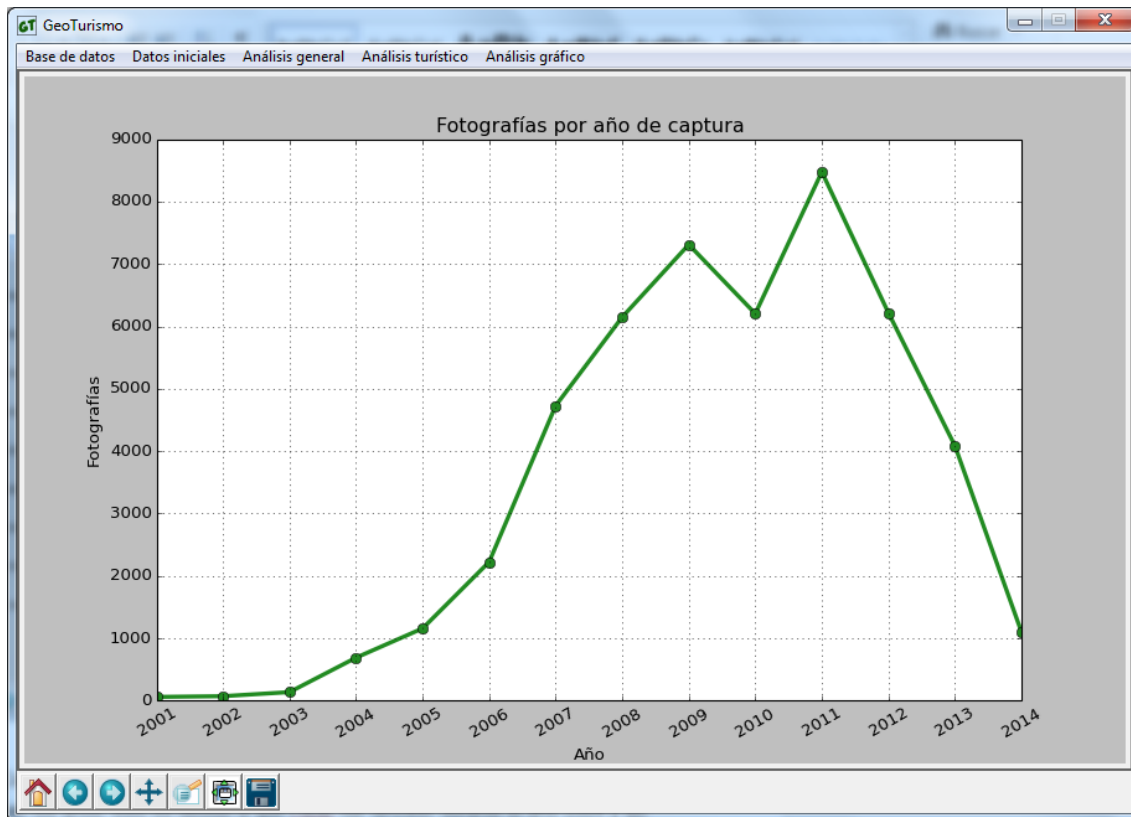


Ilustración 49: Gráfica fotografías por año de captura

Como se aprecia en la gráfica, el número de capturas va aumentando a medida que pasan los años, esto es debido a que Flickr fue lanzado durante el año 2004 y en un principio la aplicación no era conocida por el público de las redes sociales. A medida que los años pasan, la aplicación va obteniendo reconocimiento por la red hasta obtener un máximo de capturas y publicaciones en el año 2011, el punto álgido de Flickr, a partir del cual el número de publicaciones disminuye paulatinamente. Este hecho podría deberse al lanzamiento de *Instagram*, una nueva plataforma para redes sociales, aparecida en octubre 2010 que ganó adeptos rápidamente, por lo que a finales de 2014 contaba con más de 300 millones de usuarios [22].

Para la segunda consulta utilizo el mismo comando pero como parámetro pido que me devuelva el mes de captura, la consulta queda definida de la siguiente manera:

```
SELECT EXTRACT(month FROM date_taken) AS mes,
       COUNT(date_taken) AS cantidad_fotos
FROM multimedia_25830
WHERE city = 9
      AND EXTRACT (year FROM date_taken) > 2000
      AND EXTRACT (year FROM date_taken) < 2016
GROUP BY mes
ORDER BY mes ASC;
```

Obtengo una tabla con doce filas representado cada uno de los meses del año y la cantidad de fotos que se han capturado en su mes correspondiente. A continuación muestro la gráfica.

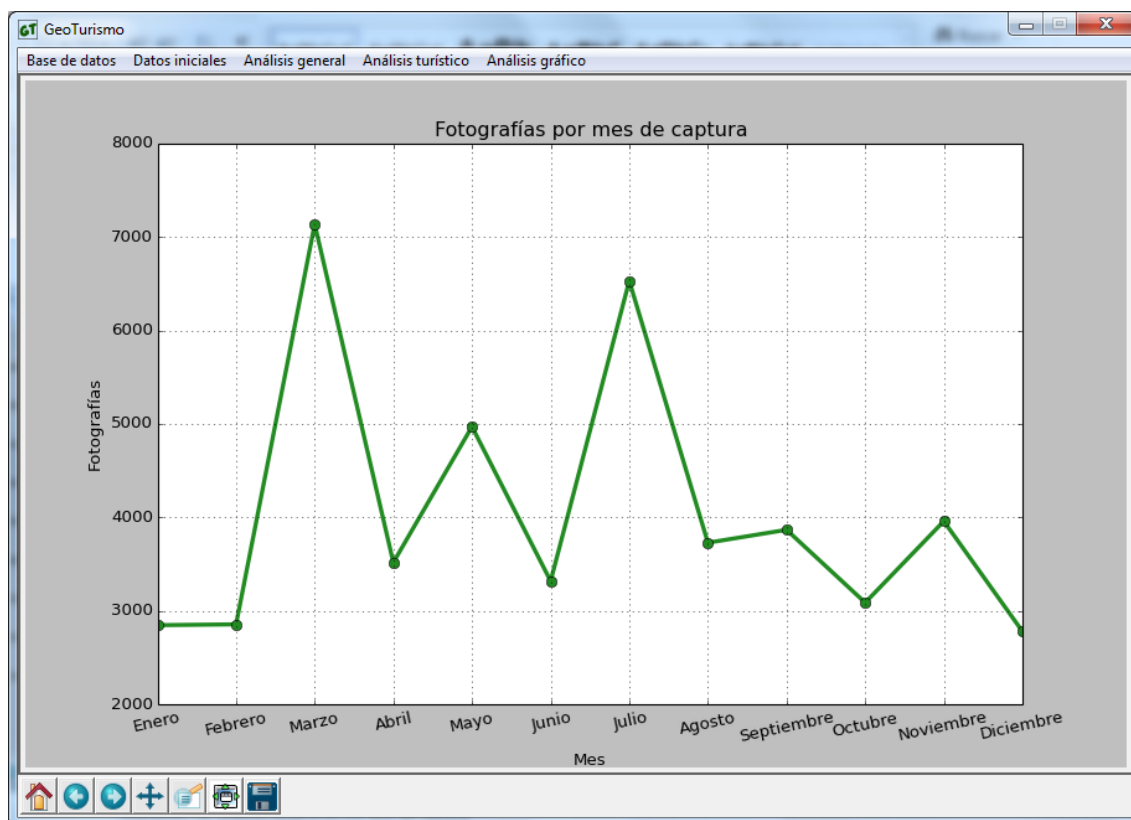


Ilustración 50: Gráfica fotografías por mes de captura

Como conclusión para esta gráfica, se aprecia un pico en el mes de Marzo, debido a las Fallas, las cuales actualmente se han convertido en un atractivo turístico muy importante y son catalogadas como una fiesta de interés turístico Internacional. El segundo de los picos de publicación de fotografías está situado en el mes de Julio, este es debido a que Valencia se sitúa en un lugar estratégico en el mar Mediterráneo con elevadas temperaturas en verano, y gracias a sus playas en plena ciudad es objetivo del turismo de verano.

La tercera consulta de intensidad de captura, se separa por días de la semana para poder comprobar cuándo los usuarios son más activos. La consulta es idéntica a las dos anteriores únicamente cambiando el comando de consulta para que nos devuelva los días de la semana. He utilizado el comando 'isodow', este ordena los días de la semana de lunes a domingo, mientras que el comando 'day' pone como primer día el domingo y esto puede causar errores en la interpretación.

```
SELECT EXTRACT(isodow FROM date_taken) AS dia,
       COUNT(date_uploaded) AS cantidad_fotos
FROM multimedia_25830
WHERE city = 9
```

```
AND EXTRACT (year FROM date_taken) > 2000
AND EXTRACT (year FROM date_taken) < 2016
GROUP BY dia
ORDER BY dia ASC;
```

La respuesta de la base de datos la proceso y creo el gráfico pertinente.

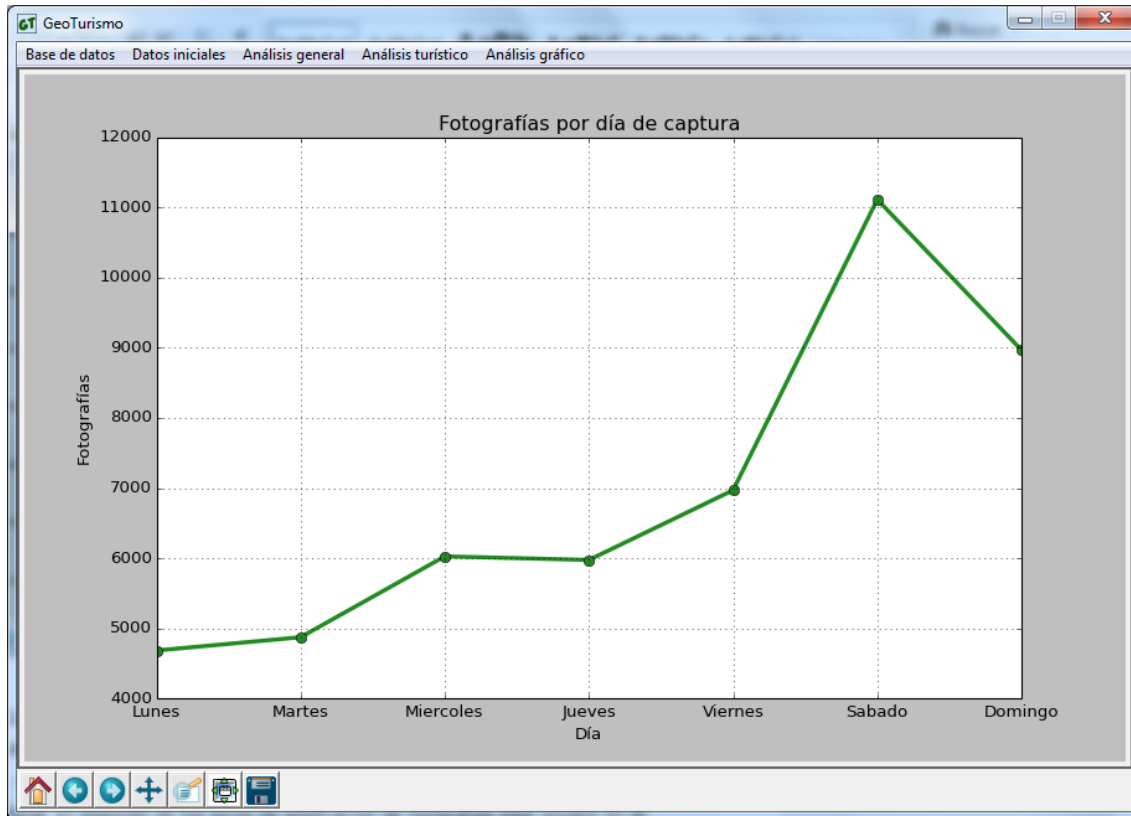


Ilustración 51: Grafica fotografías por día de captura

Se puede ver rápidamente que los días con más actividad de usuarios y por lo tanto de capturas son el fin de semana (sábado y domingo), es normal que en el fin de semana haya un pico de publicaciones ya que son días no laborables. Por último se separan las imágenes según su captura a lo largo del día. La consulta queda definida de la siguiente manera.

```
SELECT EXTRACT(hour FROM date_taken) AS hora,
COUNT(date_taken) AS cantidad_fotos
FROM multimedia_25830
WHERE city = 9
AND EXTRACT (year FROM date_taken) > 2000
AND EXTRACT (year FROM date_taken) < 2016
GROUP BY hora
ORDER BY hora ASC;
```

La solución representada mediante un gráfico es la siguiente.

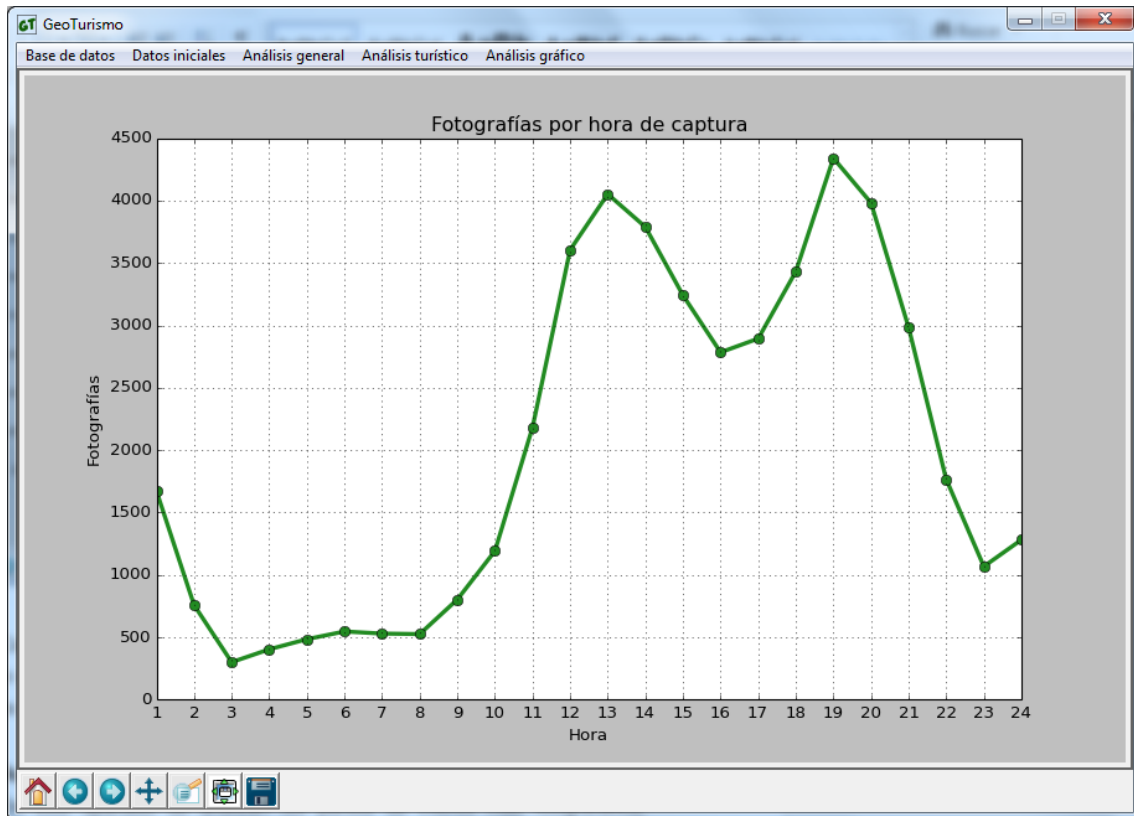


Ilustración 52: Gráfica fotografías por hora de captura

Como se puede ver, existen dos picos claramente diferenciados en los que se captura una mayor cantidad de fotografías a lo largo del día. El primero coincide con las horas centrales de la mañana y el segundo con las horas centrales de la tarde.

6.2.5. Análisis espacial

En este apartado obtengo los puntos de interés más fotografiados, considerando un punto de interés como aquel lugar de interés turístico. Para ejecutar la consulta he utilizado los polígonos de los puntos de interés descargados del catastro y les he aplicado un contorno a 50 metros, de esta manera las imágenes que se encuentren a cierta distancia del monumento también se contabilizan como pertenecientes al mismo. La consulta utiliza una función predefinida en la extensión PostGIS llamada 'DWithin', lo que realiza esta función es comprobar si por lo menos existe un punto de la segunda geometría en el interior de la primera geometría [23]. La consulta no resulta demasiado compleja pero hay que tener en cuenta los parámetros que se le introducen, su forma es la siguiente.

```
SELECT punto_interes.poi_nombre AS monumento,
       COUNT(multimedia.multimedia_identifier) AS fotos
FROM multimedia_25830, punto_interes_25830
WHERE ST_DWithin(poi_poligono, geom, 50)
GROUP BY monumento
```

ORDER BY fotos DESC;

La respuesta de la base de datos a esta consulta es una tabla extremadamente grande (86 filas), y existen monumentos que contienen muy pocas fotografías en sus inmediaciones, por lo que, para que la tabla sea más legible se han obviado los monumentos que contienen pocas fotografías o videos y se visualizan únicamente los 12 puntos de interés más visitados. Para la representación se ha utilizado un gráfico circular.

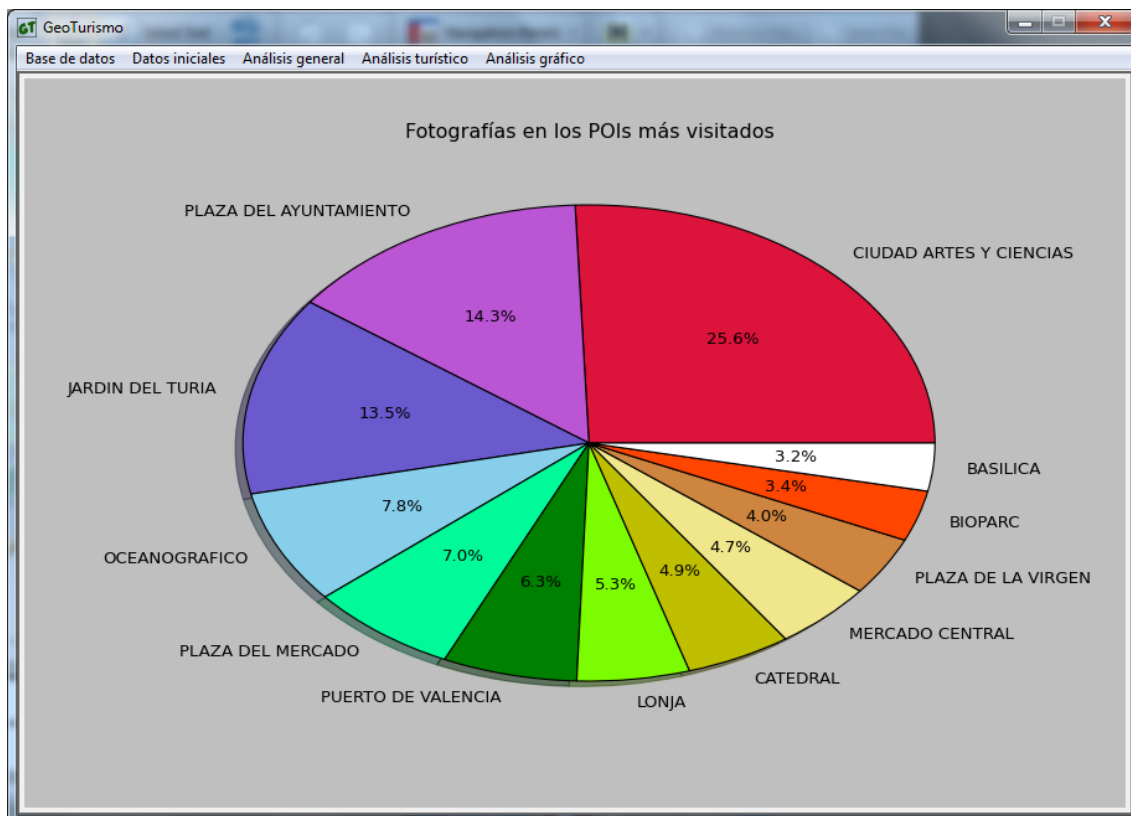


Ilustración 53: Gráfica fotografías por punto de interés

Como se observa en la gráfica, la Ciudad de las Artes y las Ciencias es el monumento de mayor interés para los usuarios que están en la ciudad de Valencia. Seguido tenemos la Plaza del Ayuntamiento y el Jardín del Turia, siguiéndoles más monumentos del centro histórico de la ciudad y otras partes del parque del río del Turia. Observando estos datos podemos concluir que la gran mayoría de las fotografías publicadas se encuentran en la zona del antiguo cauce del Turia y en algunos de los monumentos del casco histórico de la ciudad.

6.2.6. Análisis espacio-temporal

En este apartado realizo dos gráficas en las cuales agrupamos las fotografías, una por días de la semana y otra por horas del día de su momento de captura, pero al

contrario de como se hizo anteriormente, esta vez se realizara únicamente de las fotografías que se encuentren en los puntos de interés o en sus inmediaciones de los cinco monumentos más visitados. De esta manera, podré realizar una comparación de ambos resultados y ver si siguen un mismo patrón.

Las consultas sobre la base de datos son similares a las anteriormente explicadas, únicamente añadiendo el condicionante de que la geometría de las fotografías debe estar en el interior de la geometría de algún punto de interés. Se pone como ejemplo una de ellas para ver su aspecto, el cual es un poco más complejo que las anteriores.

```
SELECT COUNT(date_taken) AS cantidad_fotos
FROM multimedia_25830, punto_interes_25830
WHERE ST_DWithin(poi_poligono, geom, 50)
      AND EXTRACT(year FROM date_taken) > 2000
      AND EXTRACT(year FROM date_taken) < 2016
      AND poi_nombre = 'CIUDAD ARTES Y CIENCIAS'
GROUP BY EXTRACT(isodow FROM date_taken)
ORDER BY EXTRACT(isodow FROM date_taken) ASC;
```

A esta consulta le preceden cuatro más de la misma forma, pero cambiando el nombre del monumento. La integración de las cuatro consultas en un mismo gráfico es la siguiente:

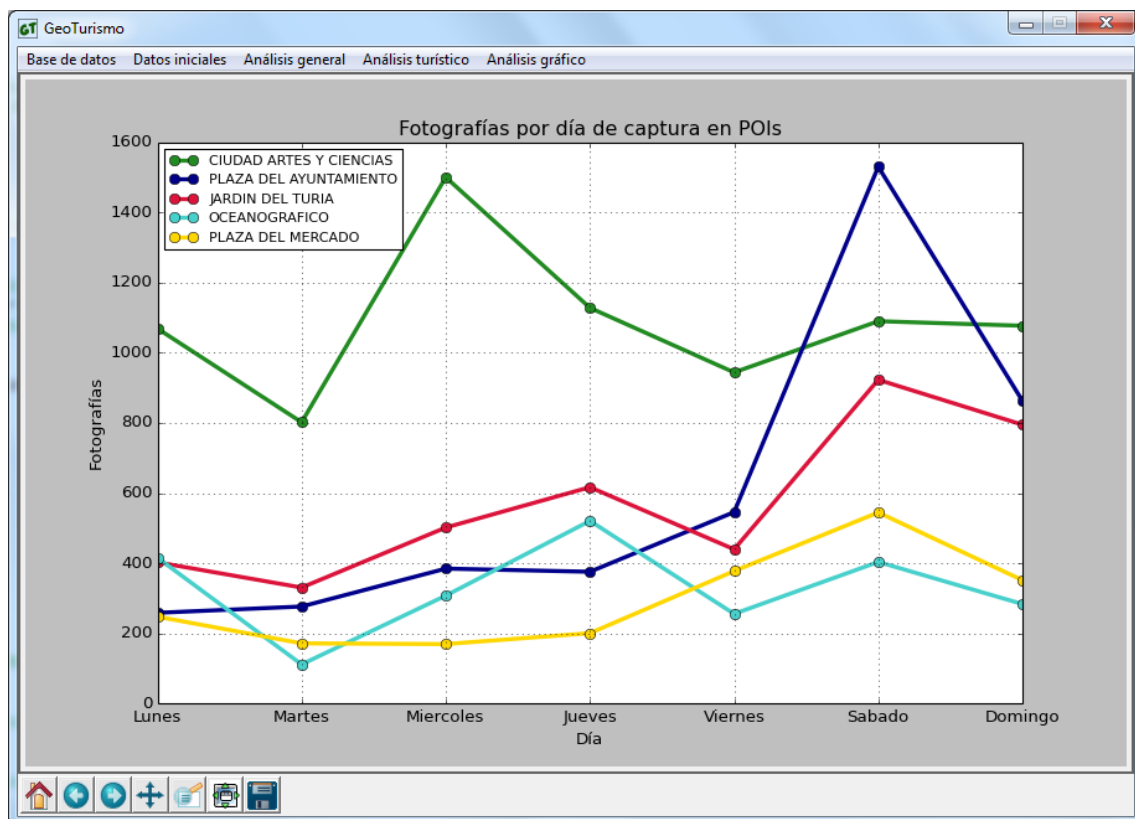


Ilustración 54: Gráfica fotografías por día de captura en POIs

Como se puede apreciar en la gráfica, todos los monumentos reciben un número bajo de visitas en los primeros días de la semana, exceptuando la Ciudad de las artes y las ciencias, la cual tiene su máximo el miércoles. Los cuatro siguientes monumentos más visitados tienen sus máximos de visitas el sábado, esto puede ser lógico, ya que el fin de semana es cuando la gente está más dispuesta a realizar visitas a los monumentos de la ciudad.

Por otra parte, se tiene un gráfico con las mismas características pero agrupando las fotos por su hora de captura durante el día, las consultas realizadas a la base de datos son idénticas a la expuesta anteriormente. La gráfica resultante es la siguiente:

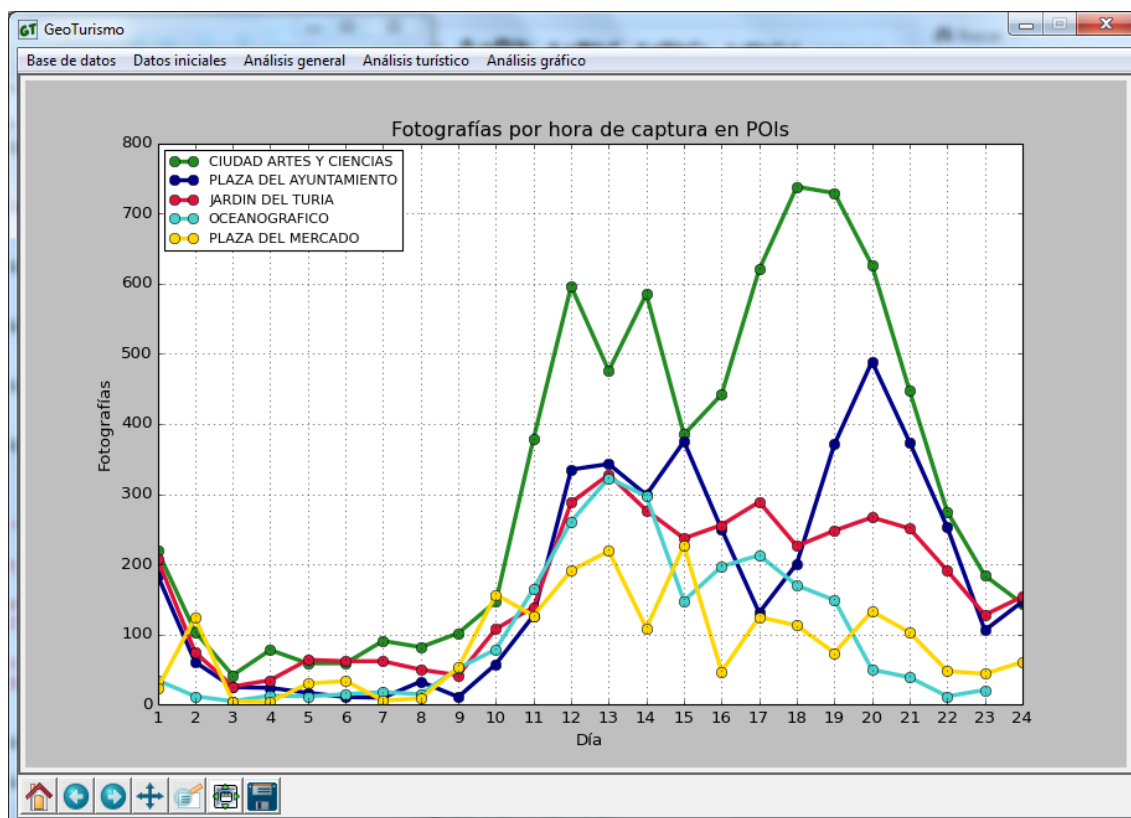


Ilustración 55: Gráfica fotografías por día de captura en Pois

En este último gráfico de la sección de usuarios totales, se puede ver de inmediato que todas las líneas tienen algo en común, y es que, en las horas de la madrugada las capturas se mantienen en mínimos. También se puede notar que en todas ellas, de manera más o menos acentuada, se aprecia un primer pico de capturas en las horas medias de la mañana. Otra de las características es que el pico máximo de captura ronda las horas de la tarde, entre las seis de la tarde y las 8 de la noche.

6.2.7. Análisis gráfico

Para el análisis gráfico he empleado la herramienta ArcMap. Con ella he creado una serie de mapas de la ciudad de Valencia. Para el cálculo de la densidad de fotografías/videos he utilizado la herramienta 'KernelDensity', la cual calcula la densidad de una entidad en su vecindad. El cálculo puede realizarse tanto para puntos como para líneas. Para el cálculo de la vecindad he establecido un radio de búsqueda de 1000 metros.

Para todo el diseño y la producción de los mapas se ha empleado la biblioteca ArcPy de Python, con la que he podido automatizar el proceso. Únicamente he tenido que crear un fichero .mxd (extensión de fichero de un documento de mapa usado por ArcMap) de partida y una serie de ficheros con la simbología a aplicar.

La serie consta de cuatro mapas ubicados en distintas zonas de Valencia, a las que he denominado: Casco histórico y alrededores, Ciudad de las Artes y las Ciencias, Puerto deportivo y playa de la Malvarrosa y Parque de Cabecera y Feria de Valencia (Benimámet). En ellos se muestra la densidad de fotografías de todos los usuarios que han publicado en la ciudad.

A continuación muestro una miniatura de ejemplo de uno de los planos de la serie:

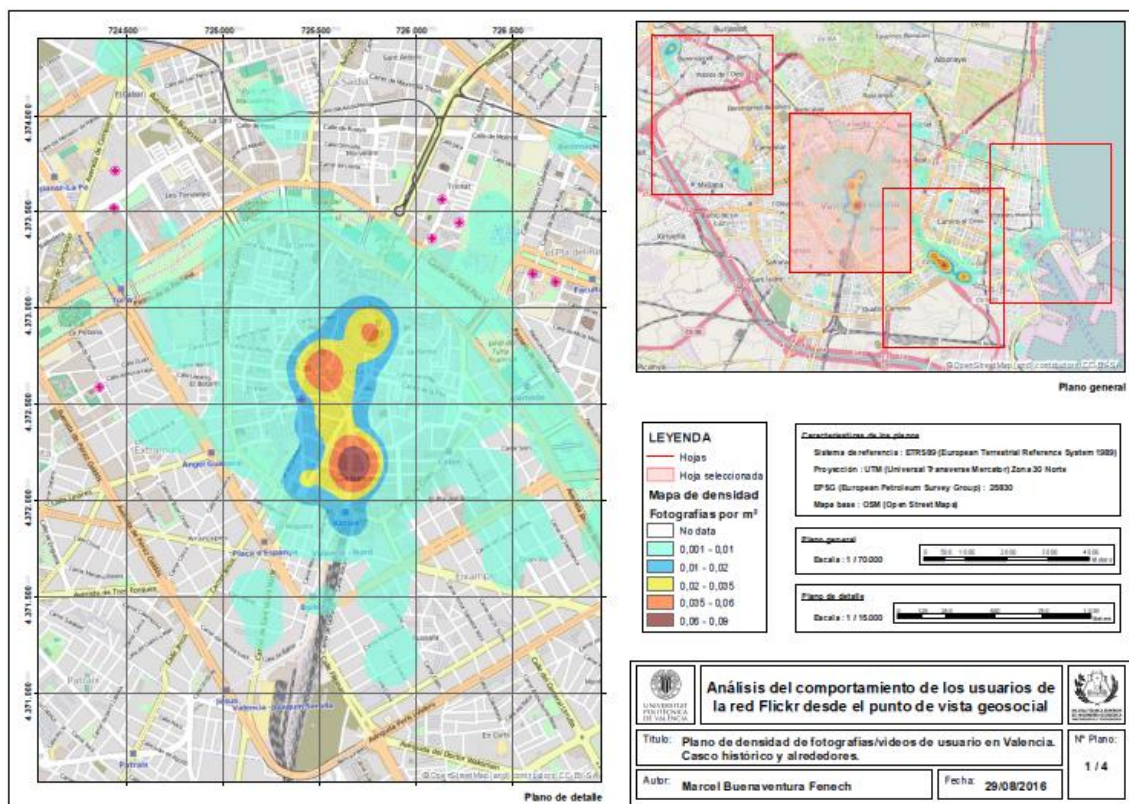


Ilustración 56: Miniatura del plano de densidad de fotografías de usuario

En la miniatura se puede ver que en la parte superior derecha, se representan los cuatro mapas de los que consta la serie, destacando con un sombreado el que se

detalla a su izquierda. Las unidades que he utilizado para la representación de la densidad son el número de fotografías por metro cuadrado. También incorporo datos sobre las escalas, sistemas de referencia y mapa base utilizado.

Para ver los planos a tamaño real ir al Anexo I - Planos de usuarios.

6.3. Análisis de turistas

En este apartado continuo con el mismo tipo de análisis y las secciones son las mismas que las del apartado anterior, la diferencia radica en que este análisis tendrá en cuenta únicamente a los usuarios que se consideran turistas dentro de la aplicación de Flickr.

Las consultas para la realización de las gráficas se omitirán en los apartados que restan del análisis por la gran similitud con las anteriores. La diferencia dentro de la consulta SQL, se encuentra en que introduzco la condición de que el atributo de 'turist' tiene que almacenar al valor 'true'. Un ejemplo de la sentencia es el siguiente:

```
SELECT COUNT (DISTINCT user_nsid)  
FROM turistas  
WHERE turist = true;
```

6.3.1. Determinación de los usuarios turistas

Este análisis posiblemente sea el más importante dentro de todo el trabajo, y para realizarlo se tienen que establecer una serie de criterios para poder discriminar si un usuarios se considera turista. Los criterios que se tuvieron en cuenta en un principio fueron los siguientes:

- Criterio 1: comprobar si un usuario ha colgado fotos en un periodo de tiempo relativamente corto.
- Criterio 2: comprobar si un usuario ha colgado fotos durante un periodo entre 4 meses y 1 año.
- Criterio 3: comprobar si un usuario ha colgado fotos cerca de los puntos de interés de la ciudad.
- Criterio 4: comprobar si un usuario ha colgado fotos en al menos 3 monumentos de la ciudad.
- Criterio 5: comprobar si un usuario ha colgado en un periodo muy corto una gran cantidad de fotos.
- Criterio 6: comprobar si la ciudad de origen del usuario es distinta de la ciudad objeto de análisis.

- Criterio 7: conocer en qué idioma escribe el usuario sus etiquetas.

A continuación paso a explicar cada uno de los criterios seguidos y la conclusión que he obtenido en su utilización.

Criterio 1:

Para evaluar este criterio he optado por un umbral de 7 días, he aplicado este umbral porque un usuario puede realizar tanto un viaje turístico 'expres' o de mayor duración. Por otro lado, tengo que aplicar un umbral inferior para poder eliminar los posibles usuarios que solo hayan utilizado la aplicación una vez y después la hayan abandonado, por este motivo establezco un umbral de 2 días como mínimo.

La consulta a la base de datos es algo más compleja que las anteriores, ya que hay que añadir condiciones más específicas. Se aplican conocimientos explicados anteriormente y queda definida de la siguiente manera.

```
SELECT DISTINCT(user_nsid), age(MAX(date_taken),MIN(date_taken))
FROM multimedia_25830
WHERE city = 9
AND EXTRACT (year FROM date_taken) > 2000
AND EXTRACT (year FROM date_taken) < 2016
GROUP BY user_nsid
HAVING extract(day FROM (MAX(date_taken)- MIN(date_taken))) <= 7
AND EXTRACT (day FROM (MAX(date_taken)- MIN(date_taken))) >= 2
ORDER BY age(MAX(date_taken),MIN(date_taken));
```

Mediante esta consulta aparecen un total de 166 usuarios que se podrían considerar turistas, pero sería demasiado precipitado establecer estos usuarios como definitivos únicamente con la aplicación de este criterio.

Criterio 2:

Para la ejecución de este criterio se establece un umbral de entre 4 meses y un año, este periodo de tiempo se ha establecido para poder obtener los posibles Erasmus que se encuentren publicando fotos en la ciudad de Valencia. Se puede decir que los Erasmus son usuarios turistas de larga duración y son tan válidos o incluso más que los turistas de corta duración. El criterio Erasmus se ha establecido ya que, España es el país que mayor cantidad de Erasmus recibe de toda Europa, y Valencia se encuentra en una de las primeras posiciones de las ciudades españolas (la Universidad de Valencia se encuentra en el puesto número 2 y la Universidad Politécnica de Valencia se encuentra en la posición número 6) [24].

La consulta es muy parecida a la del criterio anterior, únicamente tengo que añadir un intervalo en el cual contamos con 120 días de mínimo (4 meses) y un máximo de 365 días (un año). Queda formulada de la siguiente manera.

```
SELECT DISTINCT(user_nsid), extract(day FROM (MAX(date_taken)- MIN(date_taken)))  
FROM multimedia_25830  
WHERE city = 9  
AND EXTRACT (year FROM date_taken) > 2000  
AND EXTRACT (year FROM date_taken) < 2016  
GROUP BY user_nsid  
HAVING EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) <= 365  
AND EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) >= 120  
ORDER BY extract(day FROM (MAX(date_taken)- MIN(date_taken)));
```

La respuesta de esta consulta es que se consideran un total de 102 usuarios como turistas de larga duración (Erasmus).

Criterio 3:

En este caso establezco un radio de 50 metros para determinar si un usuario ha realizado capturas cerca de un punto de interés. Para ello se comprueba si un usuario tiene alguna imagen dentro del polígono del punto de interés añadiendo una distancia de 'buffer' de 50 metros. La consulta queda representada de la siguiente manera.

```
SELECT DISTINCT(user_nsid)  
FROM multimedia_25830, punto_interes_25830  
WHERE ST_DWithin(poi_poligono, geom, 50)  
GROUP BY user_nsid  
ORDER BY user_nsid;
```

De esta consulta, se concluye que un total de 1380 usuarios han publicado imágenes en las inmediaciones de un punto de interés.

Criterio 4:

Para la realización de este criterio establezco un umbral de tres monumentos visitados, he establecido este umbral, ya que un turista cuando va de viaje a alguna ciudad visita bastantes monumentos y realiza varias fotos de ellos. Para el análisis considero que un mínimo de 3 monumentos son suficientes para catalogarlo como turista. La consulta resultante es la siguiente.

```
SELECT DISTINCT(user_nsid), COUNT(DISTINCT (poi_id))  
FROM multimedia_25830, punto_interes_25830  
WHERE ST_DWithin(poi_poligono, geom, 50)  
GROUP BY user_nsid  
HAVING COUNT(DISTINCT (poi_id)) >= 3  
ORDER BY COUNT(DISTINCT (poi_id)) ASC;
```

La respuesta de esta consulta es que se consideran como turistas que han visitado por lo menos 3 monumentos un total de 585 usuarios.

Criterio 5:

Para aplicar este criterio sobre los usuarios de la base de datos, realizo un cambio en la consulta, y la adapto para obtener la fecha en la que las imágenes se colgaron en la aplicación. De esta manera, se puede saber si un usuario ha colgado una gran cantidad de fotografías en un periodo de tiempo muy corto.

Para la consulta establezco dos parámetros, el primero de ellos es que el periodo en el que se han colgado sea inferior a un día, y el segundo es que el usuario al menos debe de haber subido una cantidad de 10 fotografías o más. La consulta adquiere la siguiente forma:

```
SELECT DISTINCT(user_nsid)
FROM multimedia_25830
WHERE city = 9
GROUP BY user_nsid
HAVING extract(day FROM (MAX(date_uploaded)- MIN(date_uploaded))) <= 0
AND COUNT(multimedia_identifier)> 10
ORDER BY age(MAX(date_uploaded),MIN(date_uploaded)) ASC;
```

La respuesta de esta consulta es que un total de 215 usuarios pueden ser considerados turistas.

Criterio 6:

Para aplicar este criterio, se ha hecho uso de la API oficial de Flickr, esta se encuentra disponible para los desarrolladores que la utilicen de forma no comercial, aunque si se desea para realizar algún proyecto de tipo comercial es necesario realizar un acuerdo previo para que sea posible su uso. Para recibir una respuesta de la API en el formato deseado, es necesario seleccionar una llamada, enviar al servidor una solicitud, especificar un método y especificar los argumentos deseados. Esta API tiene muchos lenguajes de programación con los que se puede trabajar, entre ellos se puede emplear el lenguaje Python que es el que he utilizado [25].

Para poder utilizar las ventajas de esta API, se ha tenido que crear una cuenta en Flickr y registrarla como cuenta de desarrollador, de esta manera se obtiene una clave para desarrolladores y un secreto (otra clave de codificación).

Con la API se puede acceder a todo tipo de información de los usuarios, desde cuántas fotos han publicado, en qué coordenadas están situadas, sus contactos, ciudad de origen del usuario, etc. Se utiliza esta última información para poder comprobar si un usuario lo consideramos como turista o no.

Para poder utilizar la API se ha descargado una biblioteca que permite hacer peticiones, ha sido creado por un usuario desarrollador y nos permite ejecutar sentencias de forma más amigable [26]. También se ha descargado otro módulo que

permite establecer conexiones con una base de datos. A continuación se comenta parte del código que se ha utilizado para la obtención de esta información:

```
import flickrapi
import pprint
import psycopg2
```

Lo primero que he realizado es la importación de todos los módulos necesarios para poder ejecutar el código.

```
api_key = '*****'
api_secret = '*****'
flickr = flickrapi.FlickrAPI(api_key, api_secret, format='parsed-json')
conexion = "host = 'localhost' dbname = 'TFG_2' user = 'postgres' password = '*****'"
con = psycopg2.connect(conexion)
```

En la siguiente parte del código introduzco todos los parámetros necesarios para las conexiones, tanto para la conexión con la API de Flickr como con la base de datos personal. He ocultado las claves para mantener la privacidad. Como se puede ver, el formato en el que se obtiene los datos devueltos por el servidor del API es el JSON (formato de texto ligero para el intercambio de datos).

```
sentencia1 = con.cursor()
sentencia1.execute("SELECT DISTINCT(user_nsid) FROM multimedia_25830 WHERE city = 9 ORDER BY user_nsid;")
usuarios_totales = sentencia1.fetchall()
```

En esta parte del código realizo la consulta a la base de datos para obtener todos los usuarios que han publicado en la ciudad de Valencia.

```
for user in usuarios_totales:
```

```
    try:
        info = flickr.people.getInfo(user_id = user)

        location = info['person']['location']['u_content']
        nsid = info['person']['nsid']
        archi=open('C:\Users\MIMETIC\Desktop\Nuevo_documento_de_texto.txt','a')
        archi.write(user + ';' + info['person']['location']['u_content'] + '\n')
        archi.close()

        if info['person']['location']['u_content'] == '':
            archi=open('C:\Users\MIMETIC\Desktop\Nuevo_documento_de_texto.txt','a')
            archi.write(info['person']['nsid'] + ';' + 'UNKNOWN' + '\n')
            archi.close()
        else:
            archi=open('C:\Users\MIMETIC\Desktop\Nuevo_documento_de_texto.txt','a')
            archi.write(info['person']['nsid'] + ';' + info['person']['location']['u_content'] + '\n')
            archi.close()

    except:
        archi=open('C:\Users\MIMETIC\Desktop\Nuevo_documento_de_texto.txt','a')
        archi.write(user[0] + ';' + 'UNKNOWN' + '\n')
        archi.close()
```

La siguiente parte es el cuerpo del código, es en el que obtengo toda la información deseada de un usuario, por lo que, a partir de todos los usuarios obtenidos de la base de datos, se crea un bucle en el que recorreremos cada uno de ellos. En un principio, la información obtenida se encuentra en diccionarios de Python, y dentro de cada ítem de este diccionario se dividen en más diccionarios. Para acceder a una información concreta se tiene que tener un conocimiento bastante preciso de cómo están siendo devueltos los datos. Una vez se he los datos necesarios, los guardo en un documento de texto para que puedan ser leídos posteriormente. El aspecto de documento de texto es el siguiente.

```
10006319@N00;UNKNOWN  
100097189@N05;UNKNOWN  
10028574@N02;UNKNOWN  
10058483@N00;UNKNOWN  
100617112@N07;UNKNOWN
```

No he podido obtener ningún tipo de información de los usuarios, ya que casualmente ninguno de ellos tenía implementada esta información en su perfil de usuario. He realizado una prueba externa con nuestro usuario personal para comprobar si el problema es del programa, pero sí que he obtenido resultado de mi ciudad de origen.

Concluyo que este criterio para identificar a los usuarios turistas no es aplicable, ya que ninguno tiene la información necesaria en su perfil.

Criterio 7:

Para este criterio se tienen en cuenta los 'tags' de los usuarios, se tratan de una serie de etiquetas que utilizan en sus perfiles y fotografías, pero rápidamente observo que todos los usuarios utilizan varios idiomas distintos en sus 'tags', por lo que el resultado no iba a ser concluyente. Un ejemplo de 'tags' es el siguiente.

```
'cerralbo,chameleon+filter,iphone,museo,museo+cerralbo,museums'
```

Como se puede apreciar en el ejemplo anterior existen palabras escritas en dos idiomas, por lo que no se puede establecer ninguna conclusión respecto al idioma utilizado por el usuario.

Conclusión:

Una vez establecidos los criterios que se van a considerar para la obtención de los usuarios turistas, creo una consulta de todos ellos, aplicándolos de manera conjunta. La consulta queda definida de la siguiente manera:

```
SELECT DISTINCT(user_nsid)
```



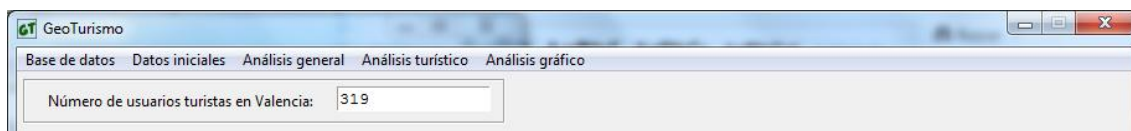
```
FROM multimedia_25830, punto_interes_25830
WHERE ST_DWithin(poi_poligono, geom, 50)
      AND EXTRACT(year FROM date_taken) > 2000
      AND EXTRACT(year FROM date_taken) < 2016
GROUP BY user_nsid
HAVING COUNT(DISTINCT (poi_id)) >= 3
      AND EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) <= 7
      AND EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) >= 2
      OR (EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) <= 365
      AND EXTRACT(day FROM (MAX(date_taken)- MIN(date_taken))) >= 120)
      OR (EXTRACT(day FROM (MAX(date_uploaded)- MIN(date_uploaded))) <= 0
      AND COUNT(multimedia_identifiser) > 10)
ORDER BY user_nsid;
```

De esta consulta obtengo como resultado que entre los usuarios que pertenecen a la aplicación de Flickr hay un total de 319 que son usuarios que han visitado Valencia por motivos turísticos.

A partir de estos resultados he creado una nueva columna en la tabla de 'multimedia_25830' en la cual almaceno la información referida a si el usuario es considerado turista o no, la nueva columna será de tipo booleano.

6.3.2. Número de turistas únicos

A continuación compruebo que la carga de todos los datos y sobre todo si la carga de la columna 'turist' de tipo booleano la he realizado correctamente. Para la comprobación realizaré una consulta sencilla para obtener el número total de usuarios turistas. El resultado obtenido mediante la aplicación es la siguiente:



El resultado es que 319 usuarios son turista, coincidente con el valor obtenido anteriormente, por lo que la carga de datos se ha realizado de manera exitosa.

6.3.3. Número de fotografías por usuario turista

En primer lugar averiguo cuántas imágenes hay en la base de datos, las cuales han sido tomadas por turistas, esta información permite obtener los estadísticos básicos de la muestra. Para obtener esta información he realizado una consulta realmente sencilla. De esta manera se obtiene un resultado de 14306 fotografías hechas por usuarios turistas.

A continuación paso a realizar la consulta que nos devuelve cuántas fotografías ha publicado cada uno de los usuarios turistas. Con esto obtengo una tabla con las

fotografías totales pertenecientes a cada usuario, que representado mediante una gráfica sale el siguiente resultado.

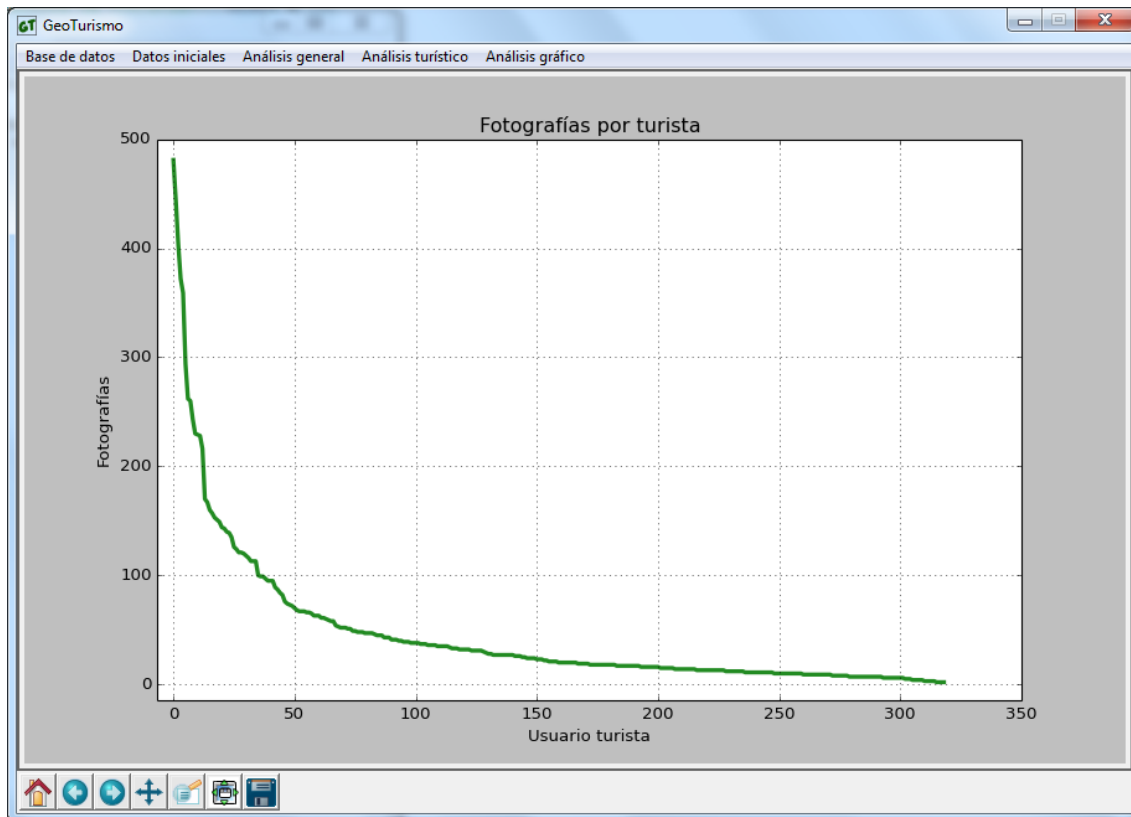


Ilustración 57: Gráfica fotografías por usuario turista

Como se puede apreciar, existen al menos 10 usuarios con una gran cantidad de fotografías publicadas, aunque la mayoría se encuentra por debajo de las 100 imágenes. La media de fotografías por usuario es de 37.85, mientras que la mediana es de 37 y tiene una desviación típica de 172.66.

6.3.4. Intervalo de captura de fotografías de turistas

En este análisis quiero obtener durante cuánto tiempo se han capturado fotografías en la ciudad de Valencia de los usuarios turistas. Para realizar este cálculo he empleado de nuevo los dos métodos descritos anteriormente (apartado 6.2.3 Intervalo de captura de fotografías).

Para el primer método, he empleado la función de PostgreSQL denominada 'AGE', se obtiene que desde la primera captura hasta la última hay un lapso de tiempo de 13 años 3 meses y 15 días. Siendo la primera captura en la fecha 1/01/2001 y la última captura a fecha de 16/04/2014.

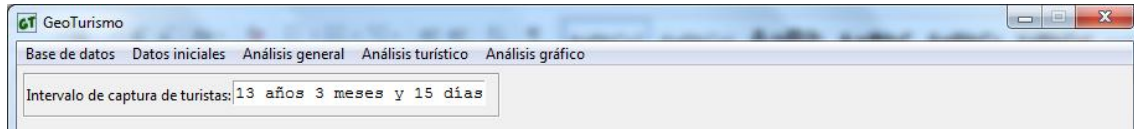


Ilustración 58: Pantalla de intervalo de captura de turistas

Para el segundo método, con el empleo de la función de PostgreSQL llamada 'EXTRACT', obtengo un resultado de 4853 días entre la fecha máxima y mínima de captura, coincidente con el resultado del anterior método.

6.3.5. Intensidad de captura de fotografías de turistas

En este apartado vuelvo a dividir el análisis en cuatro partes, la primera corresponde a la intensidad de captura separada por años, la segunda agrupando por meses del año, la tercera por días del mes y la última por horas del día.

Para la primera consulta utilizo nuevamente el comando 'EXTRACT' para poder obtener el año en el que se tomó la fotografía.

Con esta consulta obtengo una tabla con dos columnas, las cuales corresponden a los años de captura y el número total de fotografías capturadas durante ese año.

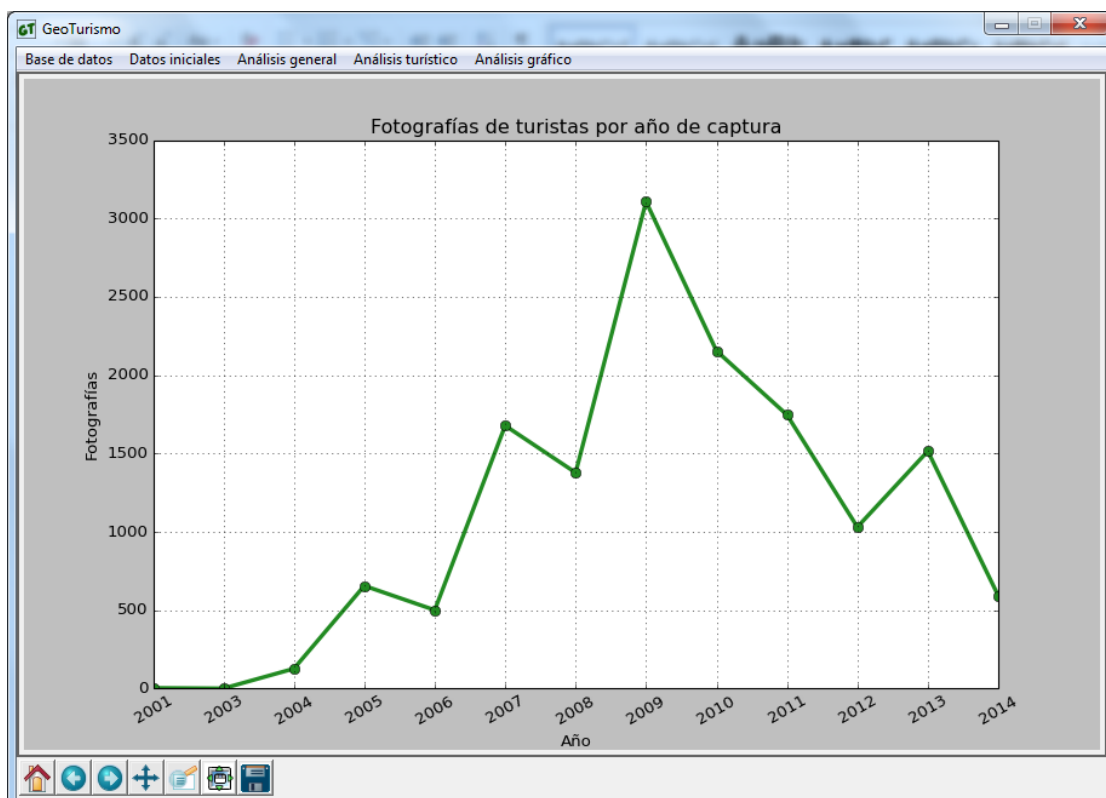


Ilustración 59: Gráfica fotografías turistas por año de captura

Se puede ver claramente que el número de fotografías aumenta de forma exponencial hasta el año 2009, siendo el año que más fotografías tiene. Después de este año, el número de capturas van disminuyendo. El año 2014 para este análisis no

lo tengo en cuenta, ya que la base de datos solo tiene fotografías hasta mitad de este año.

En segundo lugar realizo una consulta para obtener la gráfica de la distribución de fotografías de turistas por mes de captura, el resultado representado en gráfico es el siguiente.

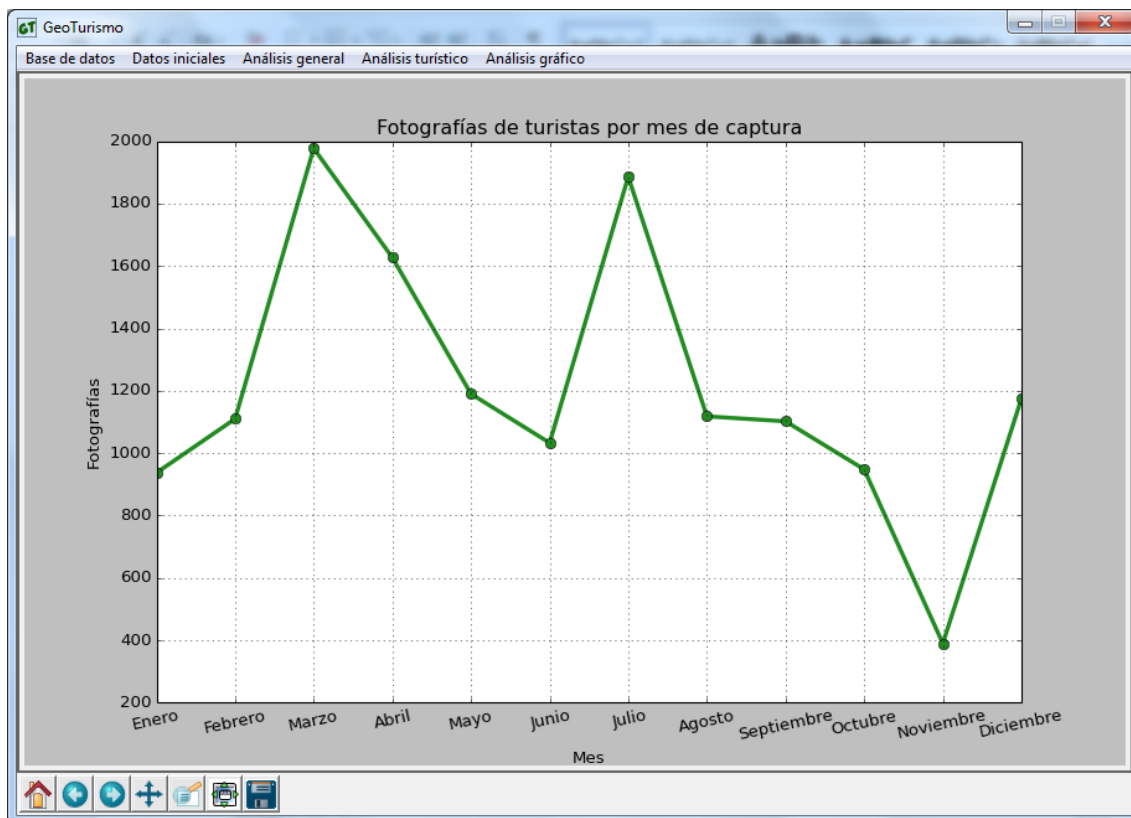


Ilustración 60: Gráfica fotografías turistas por mes de captura

El resultado de la distribución por meses, es que hay dos picos bien diferenciados, el primero corresponde al mes de marzo, es el más acusado y el que alcanza el máximo de publicaciones. El segundo pico, corresponde al mes de julio. También cabe destacar que en el mes de noviembre el número de capturas desciende hasta alcanzar el mínimo de capturas.

Para la tercera consulta de intensidad de captura, utilizo la consulta por días de la semana para poder comprobar que días de la semana los usuarios son más activos. Y su representación mediante una gráfica es la que sigue:

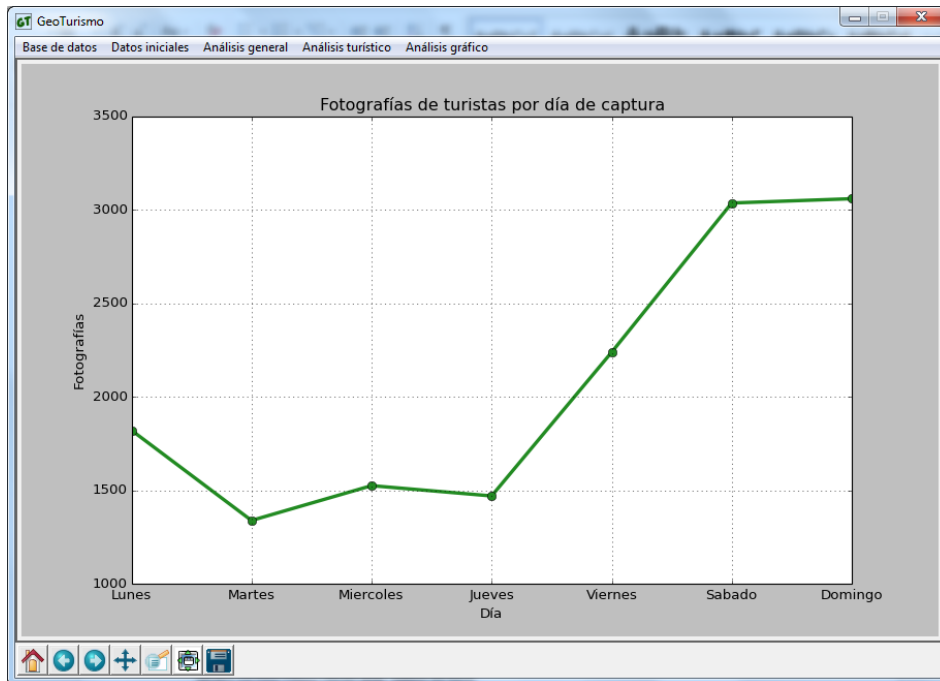


Ilustración 61: Gráfica fotografías turistas por día de captura

Como se puede observar en esta gráfica el número de capturas aumenta conforme pasan los días de la semana, teniendo sus máximos de captura en el fin de semana. También el lunes tiene un número elevado de capturas, esto podría ser debido a que las compañías 'low cost' suelen establecer los vuelos de regreso los martes y los de ida los jueves o viernes.

Por último separo las imágenes según su captura a lo largo del día. El gráfico obtenido es el siguiente:

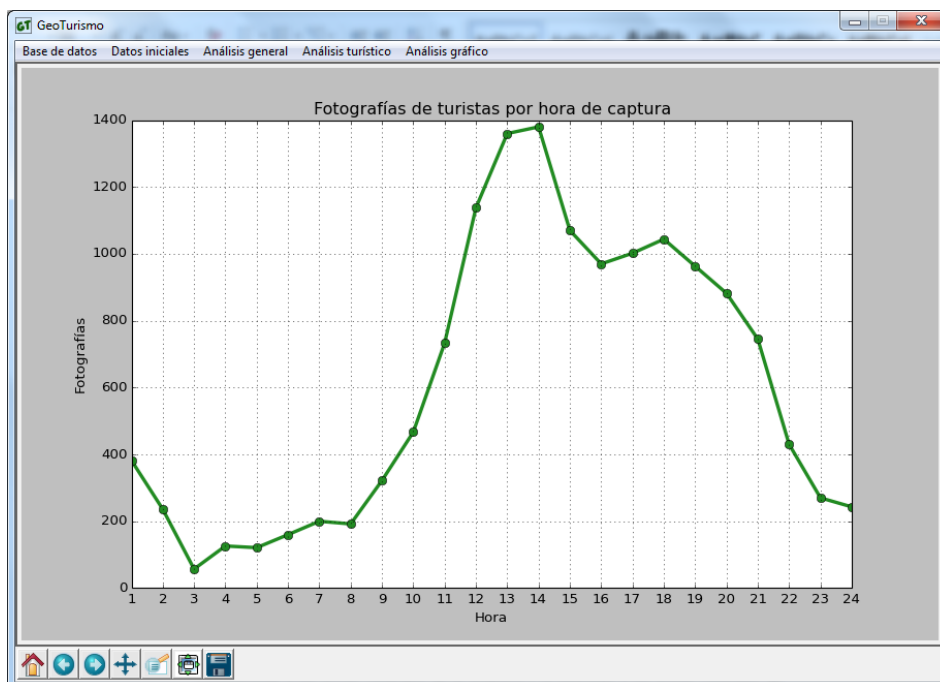


Ilustración 62: Gráfica fotografías turistas por hora de captura

Como se puede observar, tiene un único pico bien diferenciado del resto de la gráfica, corresponde a las horas centrales de la mañana. Por otro lado, en las horas centrales de la tarde se mantiene con un número de capturas elevado. El número de capturas a lo largo de la noche cae bruscamente manteniéndose en mínimos.

6.3.6. Análisis espacial de turistas

En este apartado analizo los puntos de interés más fotografiados y visitados por los usuarios turistas. Para realizar esta consulta vuelvo a establecer una distancia de buffer a los polígonos de los puntos de interés de 50 metros, esto lo realizo para considerar que las imágenes se realizan desde una distancia prudencial al monumento.

Con la consulta obtengo una tabla con dos columnas, la primera de ellas contiene el nombre de cada uno de los puntos de interés y la segunda tiene el recuento de fotografías que pertenecen al polígono del punto de interés sumando la distancia que se ha establecido como distancia prudencial de captura. A continuación muestro el gráfico para poder tener los datos más accesibles y visuales.

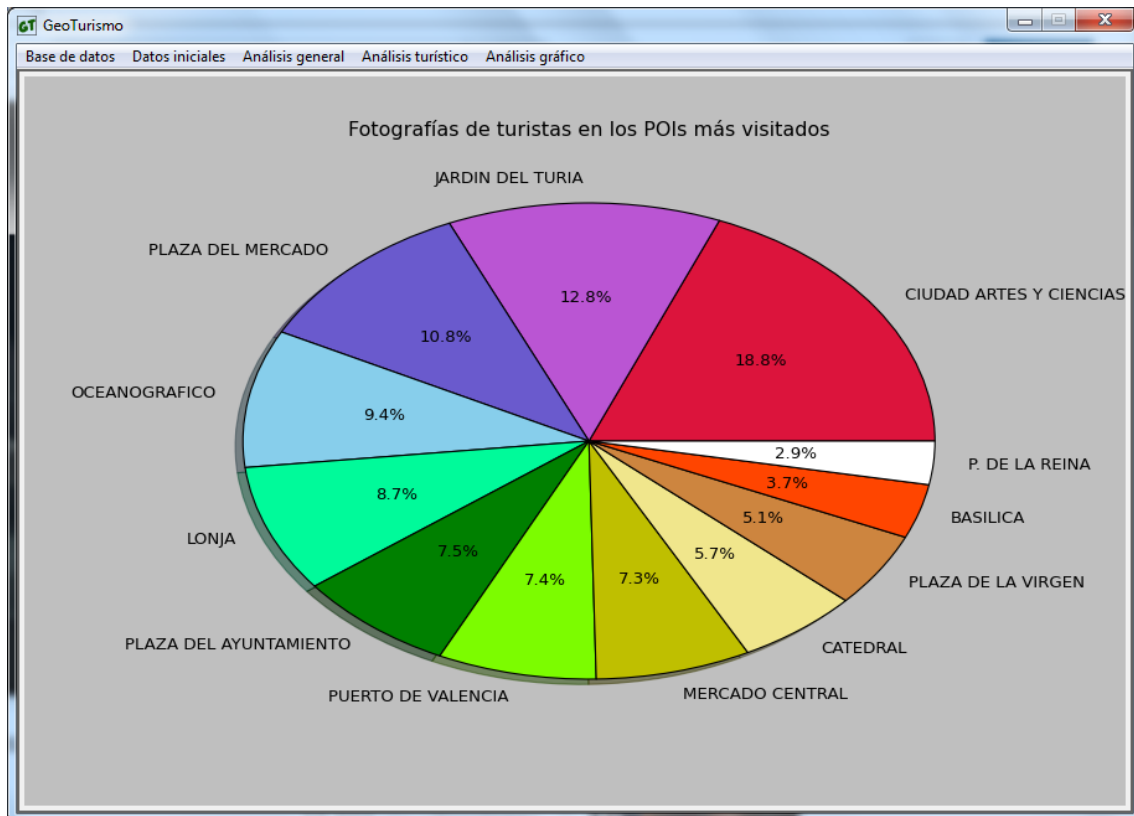


Ilustración 63: Gráfica fotografías turistas por puntos de interés

Como se puede visualizar el lugar más concurrido por los turistas se encuentra en la zona del río, más concretamente en la zona de la Ciudad de las Artes y las Ciencias y el Oceanográfico. En las zonas del casco antiguo de la ciudad, las más visitadas son

la Plaza del Ayuntamiento, el Mercado, la Lonja y la Catedral. También existe una gran atracción por el puerto de Valencia, donde se sitúa la America's Cup y alguna zona del circuito urbano de Fórmula 1.

6.3.7. Análisis turista espacio-temporal

En este apartado agrupo las imágenes por día de la semana de captura y por hora del día de captura, de esta manera podré llegar a una conclusión de en qué momentos y cuándo son más frecuentes las tomas. En este caso, filtro la base de datos para seleccionar únicamente los turistas, también selecciono solo las imágenes tomadas en las inmediaciones de los cinco puntos de interés más visitados.

Las consulta que realizo a la base de datos son idénticas a las explicadas en el mismo apartado del análisis con todos los usuarios (apartado 6.2.6 Análisis espacio-temporal).

El primer resultado de las capturas tomadas en los puntos de interés más visitados y separados por día de la semana es el siguiente:

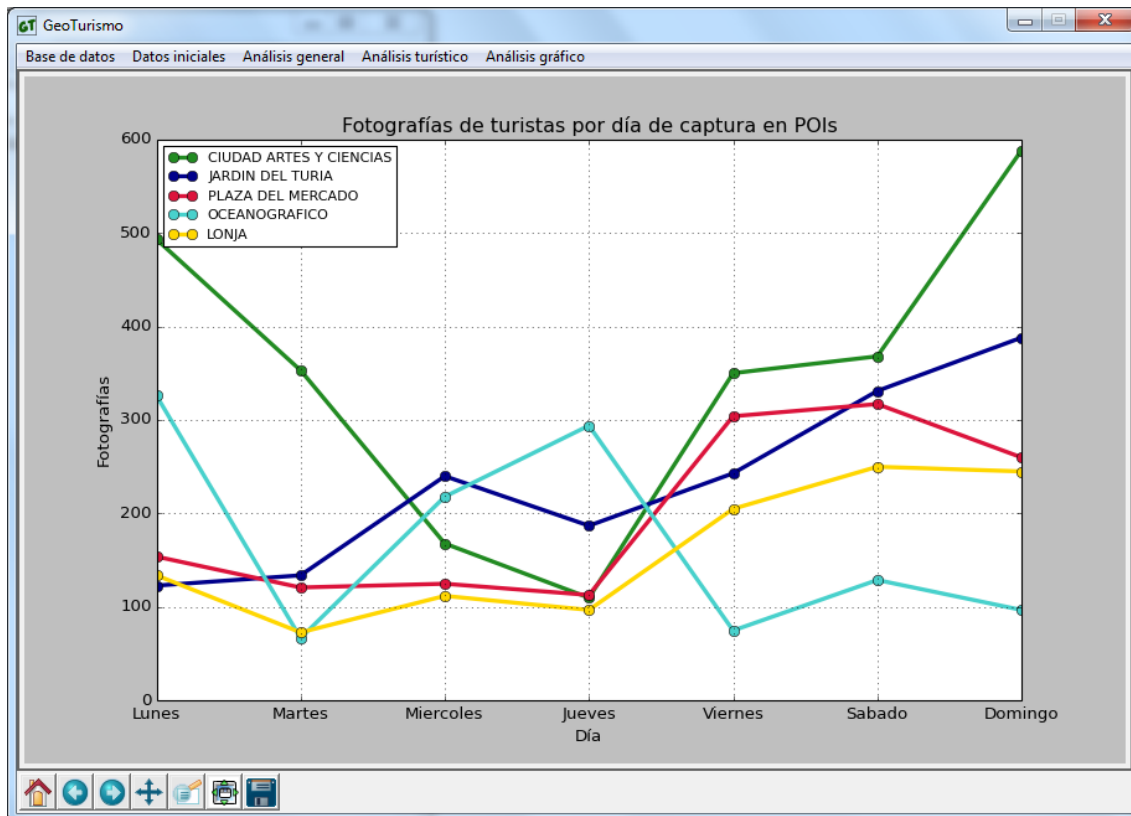


Ilustración 64: Gráfica fotografías turistas por días en POIs

Es un resultado un tanto caótico, pero se puede apreciar que todas las líneas tienden a aumentar el número de fotografías a partir del viernes, llegando a sus máximos

durante el fin de semana. Como excepción se encuentra el Oceanográfico, el cual obtiene su máximo el lunes.

A continuación presento el gráfico agrupando las capturas por la hora en la que han sido tomadas, y únicamente de los cinco monumentos más visitados de Valencia. El resultado es el siguiente:

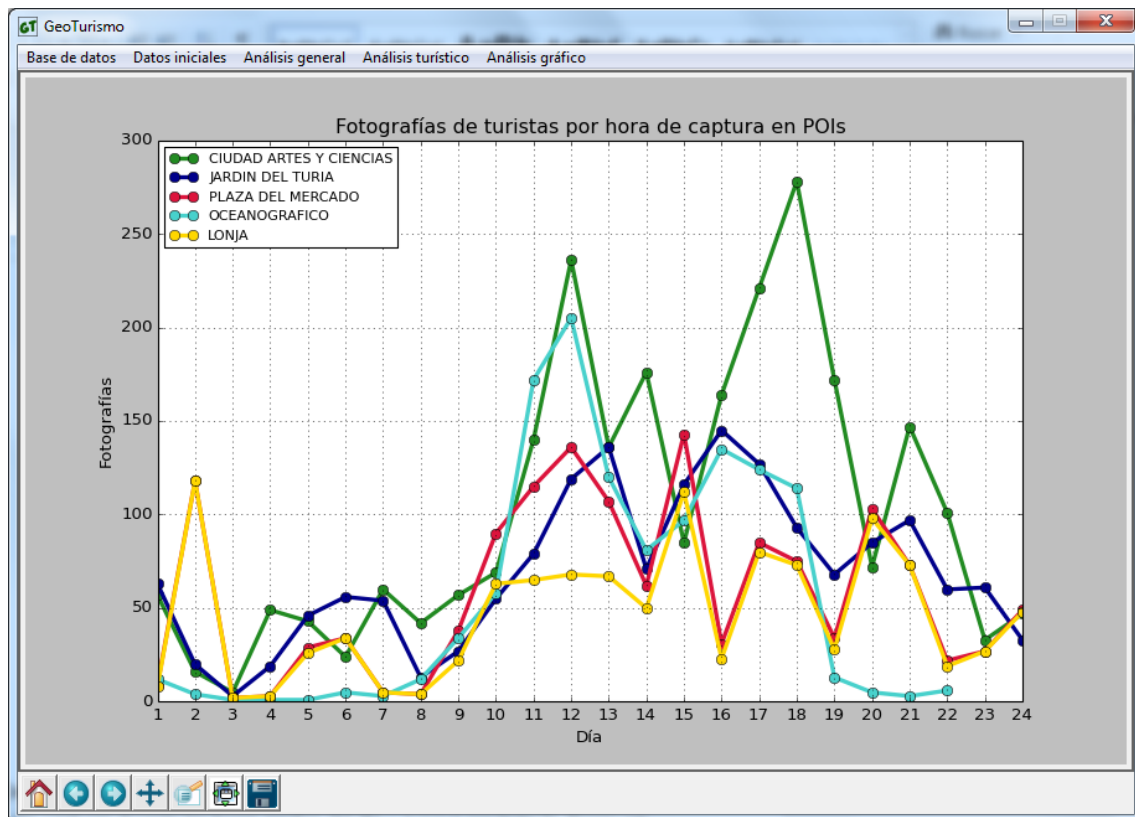


Ilustración 65: Gráfica fotografías turistas por hora en POIs

De esta representación gráfica se pueden discernir dos aumentos de capturas durante el día. El primero de ellos corresponde con las horas centrales de la mañana y el segundo durante las horas centrales de la tarde. Aunque las gráficas son un poco dispares entre ellas.

6.3.8. Análisis gráfico

El análisis gráfico que he realizado en esta sección es el mismo que el del apartado 6.2.7 únicamente cambiando los usuarios seleccionados para la representación, he cogido los usuarios turistas.

Muestro un ejemplo de esta serie de planos de turistas:

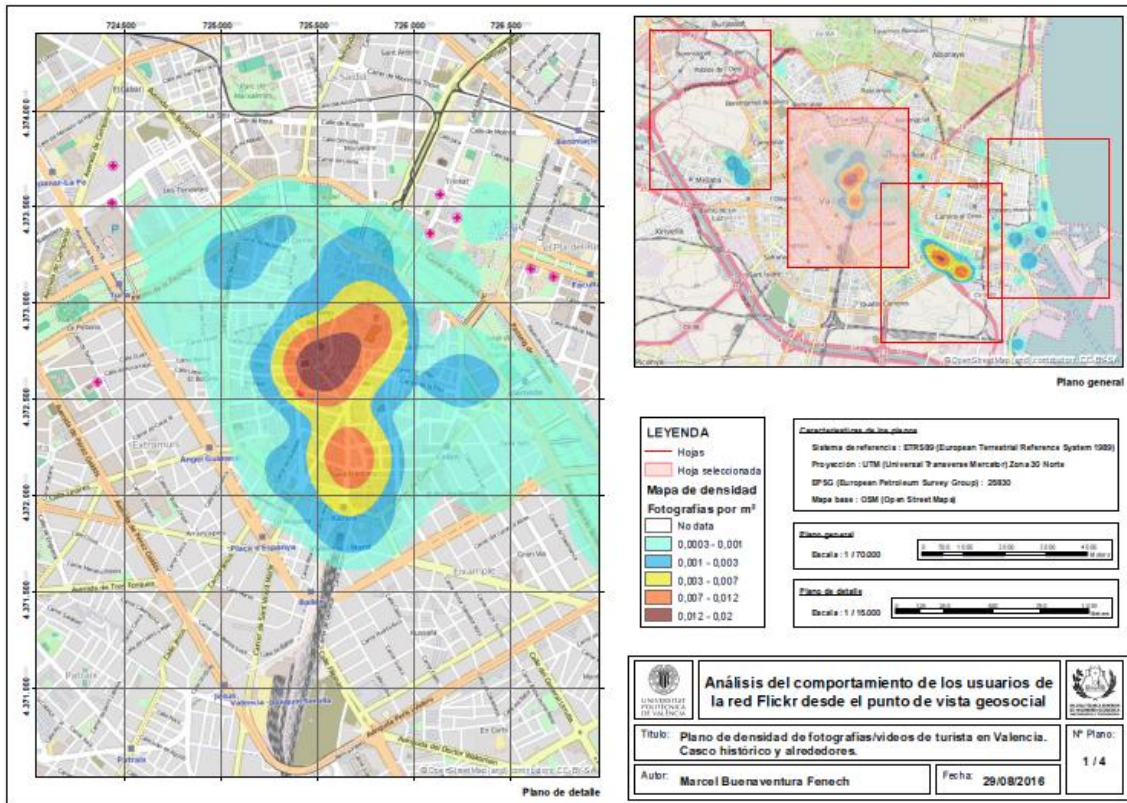


Ilustración 66: Miniatura del plano de densidad de fotografías de turistas

Para por ver toda la serie de planos ir al Anexo II – Planos de turistas.

6.4. Comparación de resultados

La última parte del análisis consistirá en realizar una comparación de los resultados del análisis de turistas con alguna página web de turismo en la ciudad de Valencia.

En primer lugar realizo la comparación con la página web de turismo de Valencia, <http://www.turisvalencia.es/>. Con este web se puede realizar una gran cantidad de consultas, buscar alojamiento en hoteles, buscar lugares típicos gastronómicos, contratar visitas guiadas, entre otros.



Ilustración 67: Ofertas de la web Turisvalencia

Lo interesante es que también se pueden buscar monumentos para visitar en la ciudad de Valencia, se pueden organizar por el tipo de emplazamiento y ajustarlo al gusto personal, pero también tiene la posibilidad de recomendarnos los 10 sitios que no debemos perdernos si realizamos una visita a la ciudad.

Entre los diez lugares que nos recomienda la web se encuentran los siguientes: Ciudad de las Artes y las Ciencias, Barrio del Carmen y sus Palacios, Plaza de la Virgen y su conjunto catedralicio, Plaza del Mercado, Lonja de la Seda, Plaza del Ayuntamiento, Marina Real y playas de Valencia, Parque Natural de la Albufera, Jardines del Turia, Bioparc y Oceanográfico.

Como se puede apreciar, todos y cada uno de los lugares recomendados por la web de turismo se encuentran en los resultados obtenidos a través de la consulta a la base de datos (ilustración 68: Gráfica fotografías turistas por puntos de interés). La única excepción es la Albufera que no la hemos incluido como punto de interés de Valencia.

Por otra parte también he encontrado otra página web de turismo de Valencia, en la cual se puede ver la distribución de fotografías de Flickr en la ciudad. La página web es la de Comunitat Valenciana (<http://comunitatvalenciana.com/>). El inconveniente es que en la representación en forma de mapa únicamente se representan 20 fotografías, por lo que para hacernos una idea de la distribución he tenido que realizar distintos zooms. A continuación mostraremos las cuatro áreas representadas en nuestros propios planos para comprobar su similitud.



Ilustración 68: Representación de fotografías de Flickr. Casco histórico

En este caso comparo la representación del Casco histórico con el Mapa de densidad de fotografías de turistas (ver anexo II – Planos de turistas, Plano 1/4). A simple vista se puede ver que las dos representaciones tienen un gran parecido en cuanto a la situación de los puntos con mayor densidad.



Ilustración 69: Representación de fotografías de Flickr. Artes y Ciencias

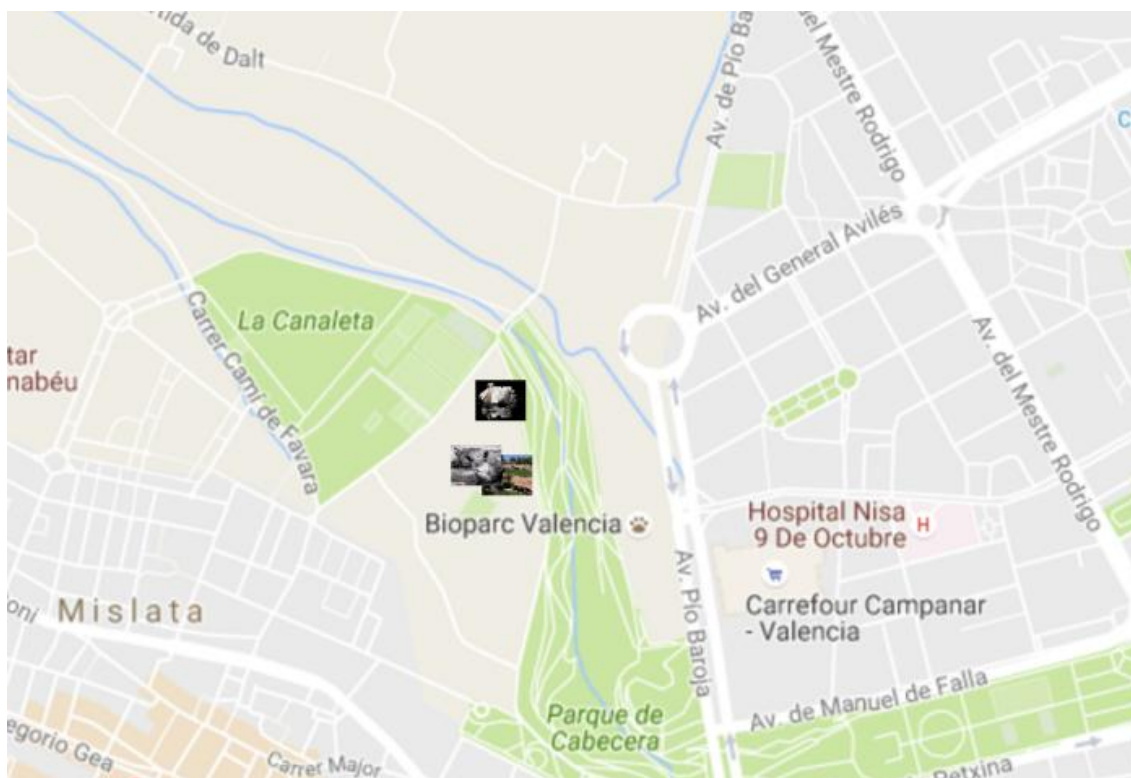


Ilustración 71: Representación de fotografías de Flickr. Parque cabecera.

En este caso, la ilustración 72 también representa el mismo resultado que el mapa creado a partir de la base de datos (ver anexo II – Planos de turistas, Plano 4/4). Aunque la información ofrecida por la página web en este emplazamiento es un poco escasa.

7. CONCLUSIONES

Para concluir con la memoria, expongo las principales conclusiones del proyecto y las conclusiones personales a las que he llegado. También mencionaré posibles líneas de trabajo, que podrían mejorar la experiencia del usuario con la aplicación.

7.1. Conclusiones

Las redes sociales cada vez están más generalizadas, se encuentran en cualquier lugar y forman parte de la vida cotidiana. Por ello, se generan y almacenan gran cantidad de datos a diario. Estos datos son un recurso extremadamente valorado y tanto empresas públicas como privadas se quieren beneficiar de sus posibilidades.

El uso masivo de las redes sociales junto con el avance de la tecnología, ha hecho posible obtener datos de calidad, en el caso de mi estudio, su importancia radica en la calidad de las medidas realizadas por el GPS de los dispositivos móviles, esto ha hecho posible este estudio, ya que proporciona coordenadas sobre el terreno con gran precisión.

Personalmente, pienso que el proyecto realizado ha sido de gran ayuda para mi formación académica. He tenido que refrescar conocimientos de los primeros cursos de Grado y darle un enfoque más práctico.

Debo decir que el proyecto ha requerido un gran esfuerzo y muchas horas de dedicación, ya que en él se tratan temas y se emplean tecnologías y herramientas que no había utilizado durante la trayectoria del Grado.

Entre las dificultades que me han surgido a lo largo de la elaboración del trabajo, cabe destacar tres: la primera ha sido desarrollar una aplicación programada en Python, nunca había programado en este lenguaje y he tenido que buscar muchísima información a través de libros e internet durante mi estudio. Para mí, aprender este lenguaje se ha convertido casi en una obsesión, ya que me he dado cuenta de la gran importancia que conlleva la automatización de los procesos a la hora de reducir costes y de optimizar el tiempo de trabajo. Otro de los problemas, ha sido tratar con una base de datos tan grande, ya que contaba con un total de 540.000 fotografías con sus distintos atributos, dada la escasa potencia de mi ordenador personal, he tenido que realizar modificaciones para que pudiese soportar ciertas consultas. Por último, los mapas de densidad, automatizar el proceso de creación ha sido extremadamente costoso y duro. He tenido que utilizar una librería muy complicada, con mucha información a través de internet, de la web oficial de ArcGis, y del libro “ArcPy and ArcGIS – Geospatial Analysis with Python”.

7.2. Trabajo futuro

El proyecto ha cumplido con todos los objetivos propuestos, y aunque han ido surgiendo distintas dificultades durante el camino, estas se han ido solventando lo mejor posible. No obstante, como todo trabajo, siempre se pueden realizar mejoras en ciertos aspectos o complementar el análisis con más resultados.

Una de estas posibilidades, es realizar planos en los que se muestren las rutas turísticas de los usuarios más activos. Habría que ordenar las fotos de cada usuario y realizar una ruta lógica por los monumentos que ha visitado. Aunque ya había comenzado a trabajar en ello y tenía un análisis bastante completo, no he podido incluirlo en el trabajo, ya que no me daba tiempo a elaborar las conclusiones finales.

También se podrían prever puntos de interés, considerando la cantidad de fotografías que se encuentren en un determinado lugar, y a posteriori considerar si se podría incluir en el catálogo de puntos de interés turístico.

Por último, añadir que como trabajo a largo plazo, se podría crear una aplicación que contase con los millones de fotografías de toda la base de datos de Flickr, que permitiese realizar el análisis para cualquier ciudad del mundo que se encontrase en la base de datos.

8. Referencias bibliográficas

[1] Sistemas de Información Geográfica.

<<http://www.almazaningenieros.es/data/archivo/Sistemas%20de%20Informacion%20Geografica.pdf>>

[2] PostgreSQL. Características, limitaciones y ventajas.

<<http://postgresql-dbms.blogspot.com.es/p/limitaciones-puntos-de-recuperacion.html>>

[3] Download Python.

<<https://www.python.org/downloads/>> (18/08/2016)

[4] Creative Commons.

<<https://creativecommons.org/>>

[5] AWS. Cloud Computing.

<<https://aws.amazon.com/es/>>

[6] Well Known Text. OGC.

<<http://www.opengeospatial.org/standards/wkt-crs>>

[7] Creative Commons. Atribución 4.0 Internacional (CC BY 4.0).

<<https://creativecommons.org/licenses/by/4.0/deed.es>>

[8] Sede Electrónica del Catastro.

<<http://www.sedecatastro.gob.es/>>

[9] ST_Transform. PostGis.

<http://postgis.org/docs/ST_Transform.html>

[10] 24.1. Tkinter - Python interface to Tcl/Tk - Python 2.7.12 documentation.

<<https://docs.python.org/2/library/tkinter.html>>

[11] 8.18. pprint - Data pretty printer - Python 2.7.12 documentation.

<<https://docs.python.org/2/library/pprint.html>>

[12] PostgreSQL + Python. Psycopg.

<<http://initd.org/psycopg/>>

[13] ¿Qué es Arcpy? ArcGis for Desktop.

<<http://desktop.arcgis.com/es/arcmap/10.3/analyze/arcpy/what-is-arcpy-.htm>>

[14] Numpy.

<<http://www.numpy.org/>>

[15] Matplotlib: python plotting. Matplotlib 1.5.1 documentation.

<<http://matplotlib.org/>>

[16] PIL 1.1.6: Python Package Index.

<<https://pypi.python.org/pypi/PIL>>

[17] 15.3. time - Time access and conversions - Python 2.7.12 documentation.
<<https://docs.python.org/2/library/time.html>>

[18] INE. Instituto Nacional de Estadística.
<<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft11%2Fp16028&file=inebase&L=0>> [Consulta: 6 de Junio de 2016]

[19] Mi portal financiero.
<<http://www.miportalfinanciero.es/articulo/ciudades-mas-turisticas-de-espana/>> [Consulta: 6 de Junio de 2016]

[20] PostgreSQL. Documentation 8.4 Date/Time functions and operators
<<https://www.postgresql.org/docs/8.4/static/functions-datetime.html>>
[Consulta: 20 de Mayo de 2016]

[21] PostgreSQL. Documentation 8.4 Date/Time functions and operators
<<https://www.postgresql.org/docs/8.4/static/functions-datetime.html#FUNCTIONS-DATETIME-EXTRACT>> [Consulta: 21 de Mayo de 2016]

[22] Instagram. Wikipedia.
<<https://es.wikipedia.org/wiki/Instagram>> [Consulta: 14 de Junio de 2016]

[23] ST_DWithin. PostGIS.
<http://postgis.net/docs/ST_DWithin.html> [Consulta: 25 de Mayo de 2016]

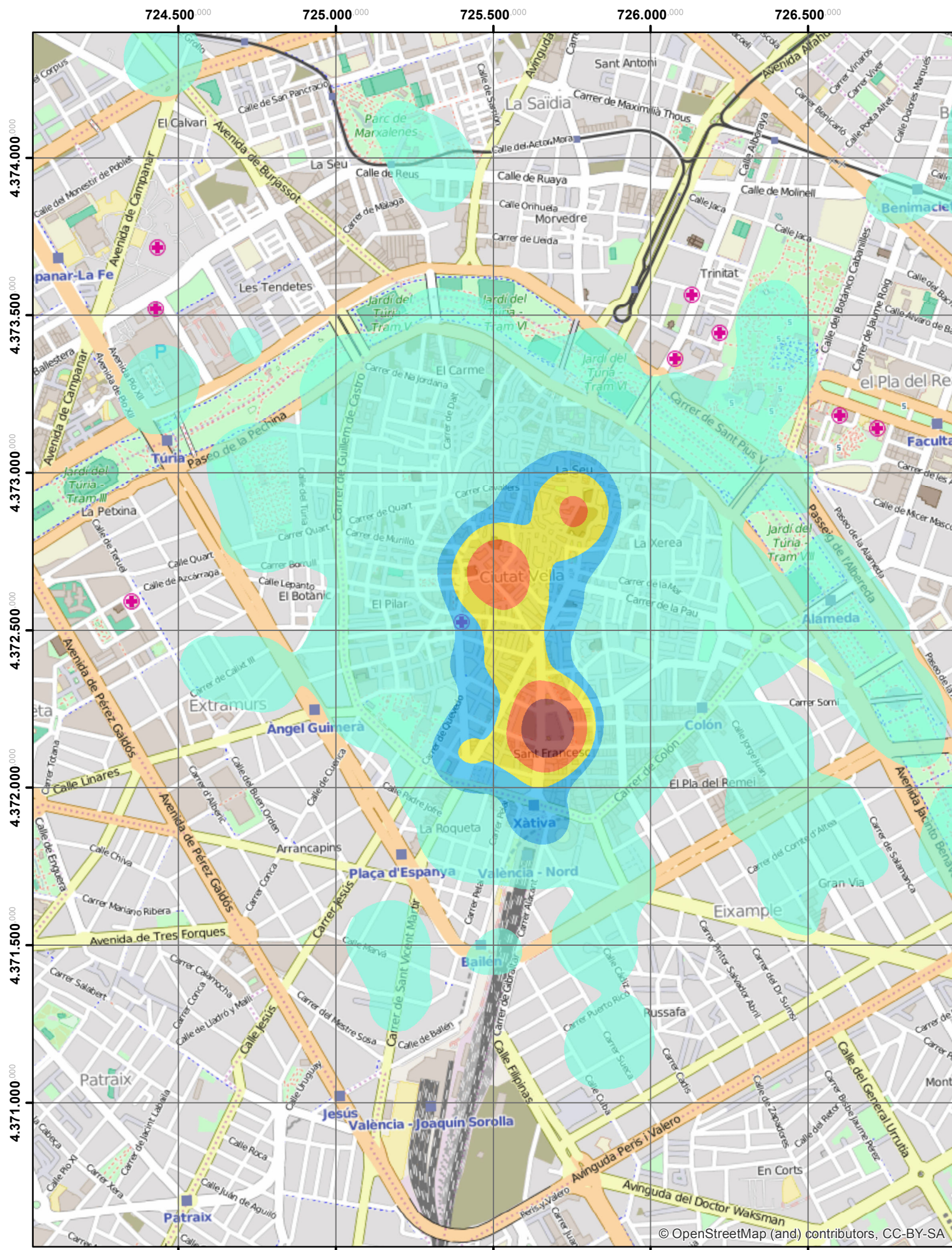
[24] La Vanguardia.
<<http://www.lavanguardia.com/vangdata/20150923/54436759424/destinos-erasmus-mas-populares.html>> [Consulta: 28 de Mayo de 2016]

[25] Servicio de Flickr. API Flickr.
<<https://www.flickr.com/services/api/>> [Consulta: 5 de Junio de 2016]

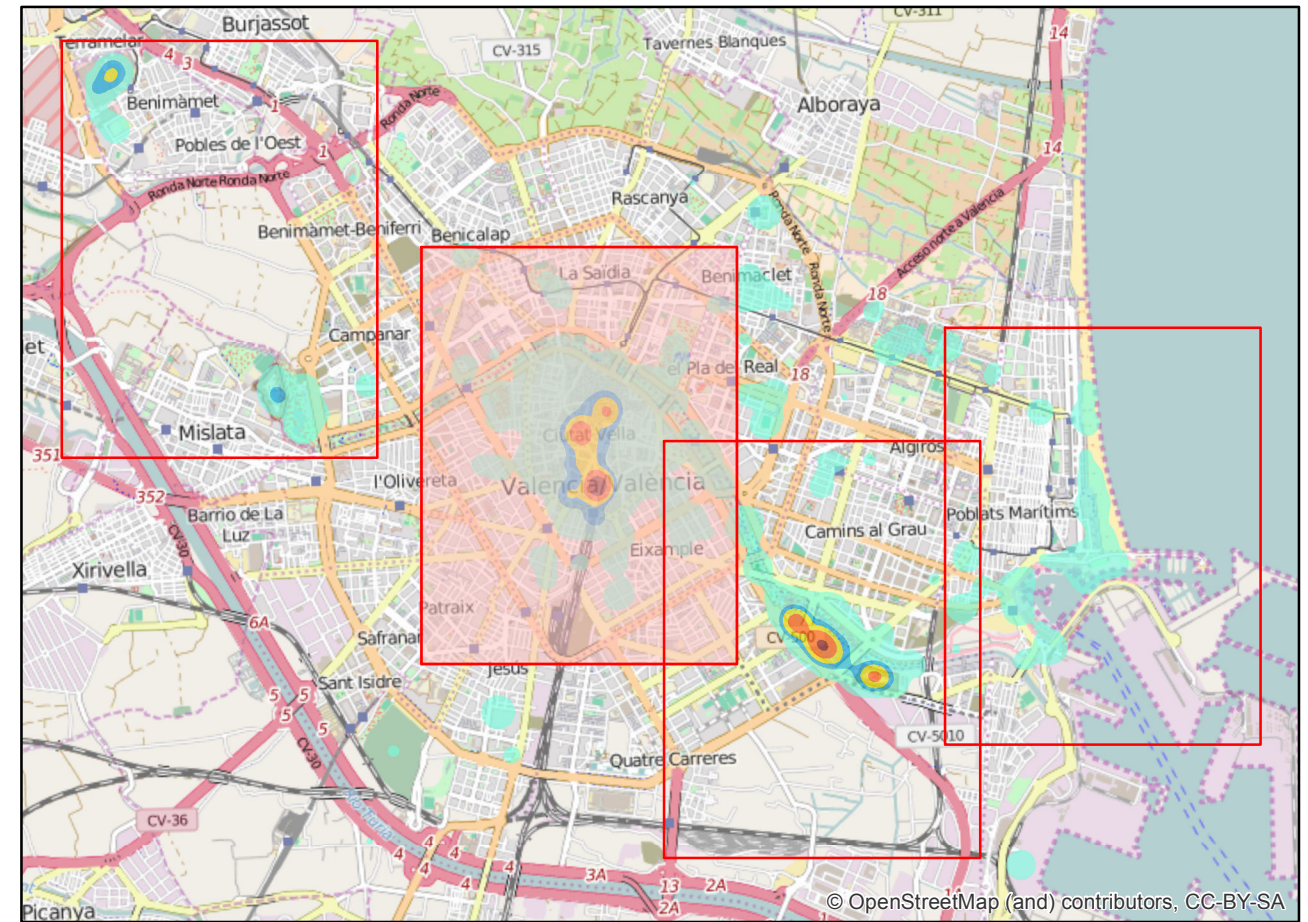
[26] Python Flickr API kit -Stüvel Stuff.
<<https://stuvel.eu/flickrapi>> [Consulta: 5 de Junio de 2016]

ANEXO I – Planos de usuarios

1. Plano de densidad de fotografías/videos de usuarios en Valencia. Casco histórico y alrededores.



Plano de detalle



Plano general

LEYENDA

- Hojas
 - Hoja seleccionada
- Mapa de densidad**
- Fotografías por m²**
- No data
 - 0,001 - 0,01
 - 0,01 - 0,02
 - 0,02 - 0,035
 - 0,035 - 0,06
 - 0,06 - 0,09

Características de los planos

Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)

Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte

EPSG (European Petroleum Survey Group) : 25830

Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



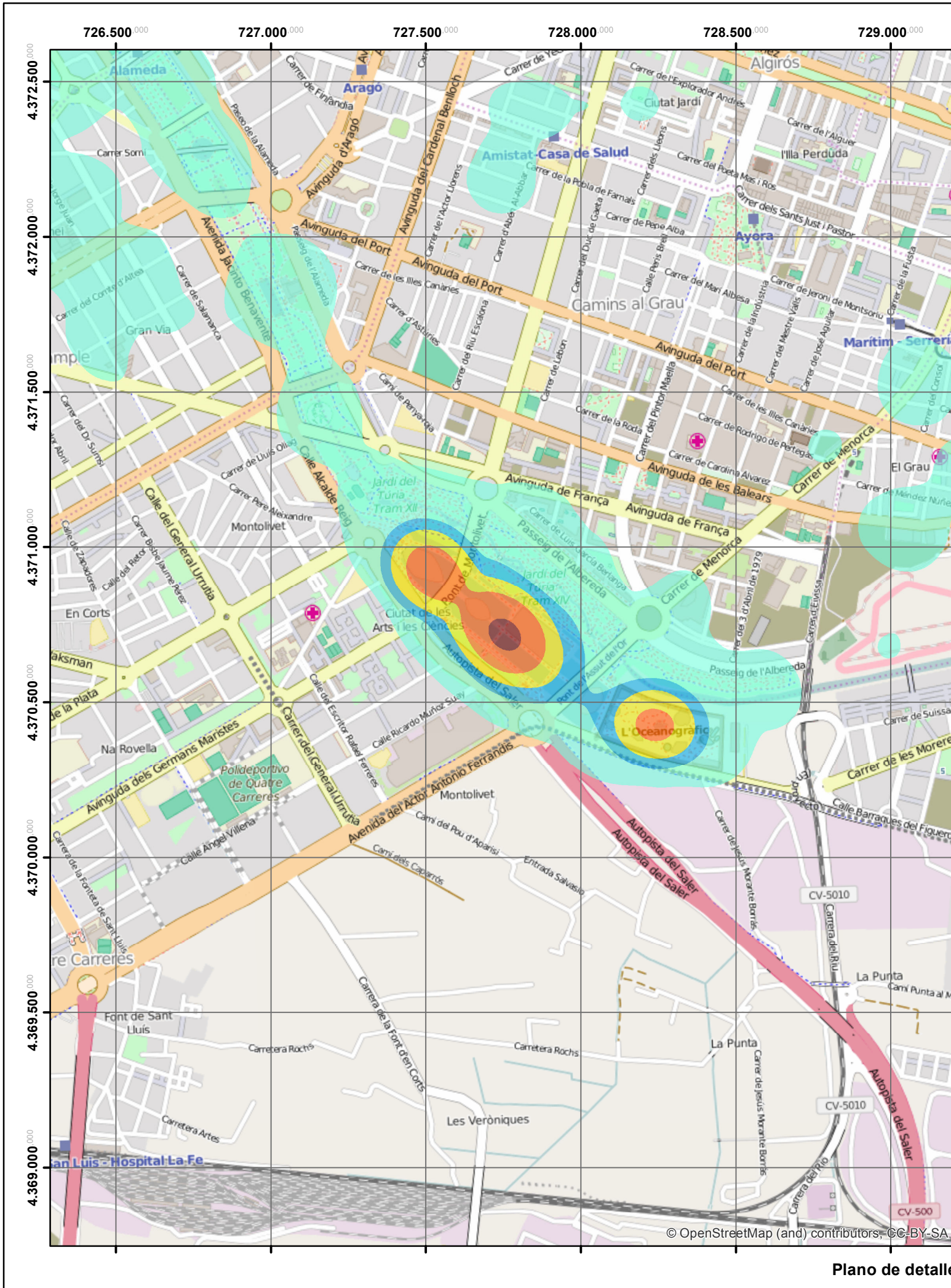
Título: **Plano de densidad de fotografías/videos de usuario en Valencia. Casco histórico y alrededores.**

Nº Plano: **1 / 4**

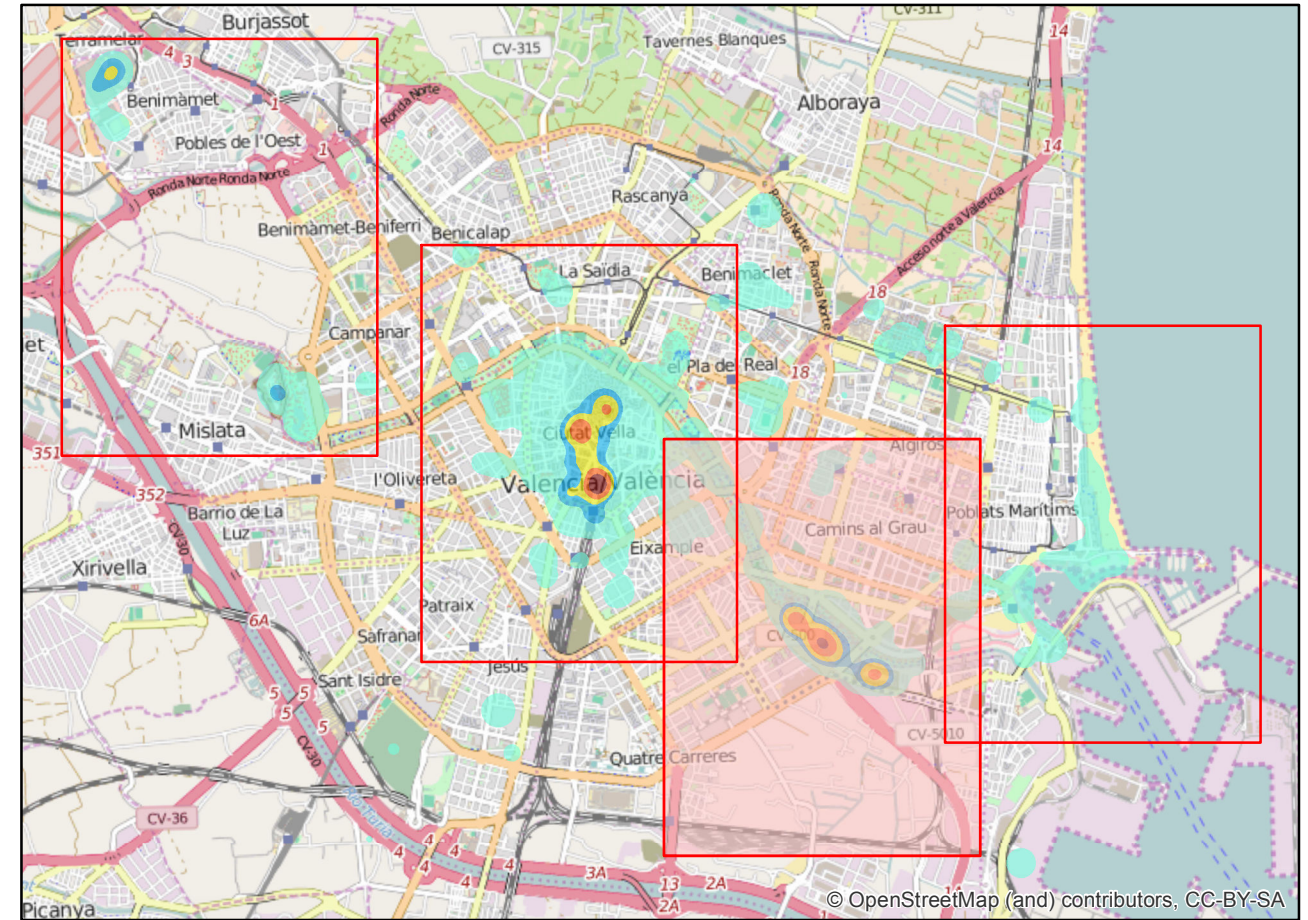
Autor: **Marcel Buenaventura Fenech**

Fecha: **29/08/2016**

2. Plano de densidad de fotografías/videos de usuarios en Valencia. Ciudad de las Artes y las Ciencias.



Plano de detalle



Plano general

LEYENDA

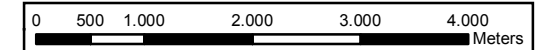
- Hojas
 - Hoja seleccionada
- Mapa de densidad**
- Fotografías por m²**
- No data
 - 0,001 - 0,01
 - 0,01 - 0,02
 - 0,02 - 0,035
 - 0,035 - 0,06
 - 0,06 - 0,09

Características de los planos

Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)
 Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte
 EPSG (European Petroleum Survey Group) : 25830
 Mapa base : OSM (Open Street Maps)

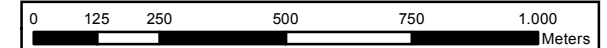
Plano general

Escala : 1 / 70.000



Plano de detalle

Escala : 1 / 15.000



UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA GEODÉSICA CARTOGRÁFICA Y TOPOGRÁFICA

Título: **Plano de densidad de fotografías/videos de usuario en Valencia. Ciudad de las Artes y las Ciencias.**

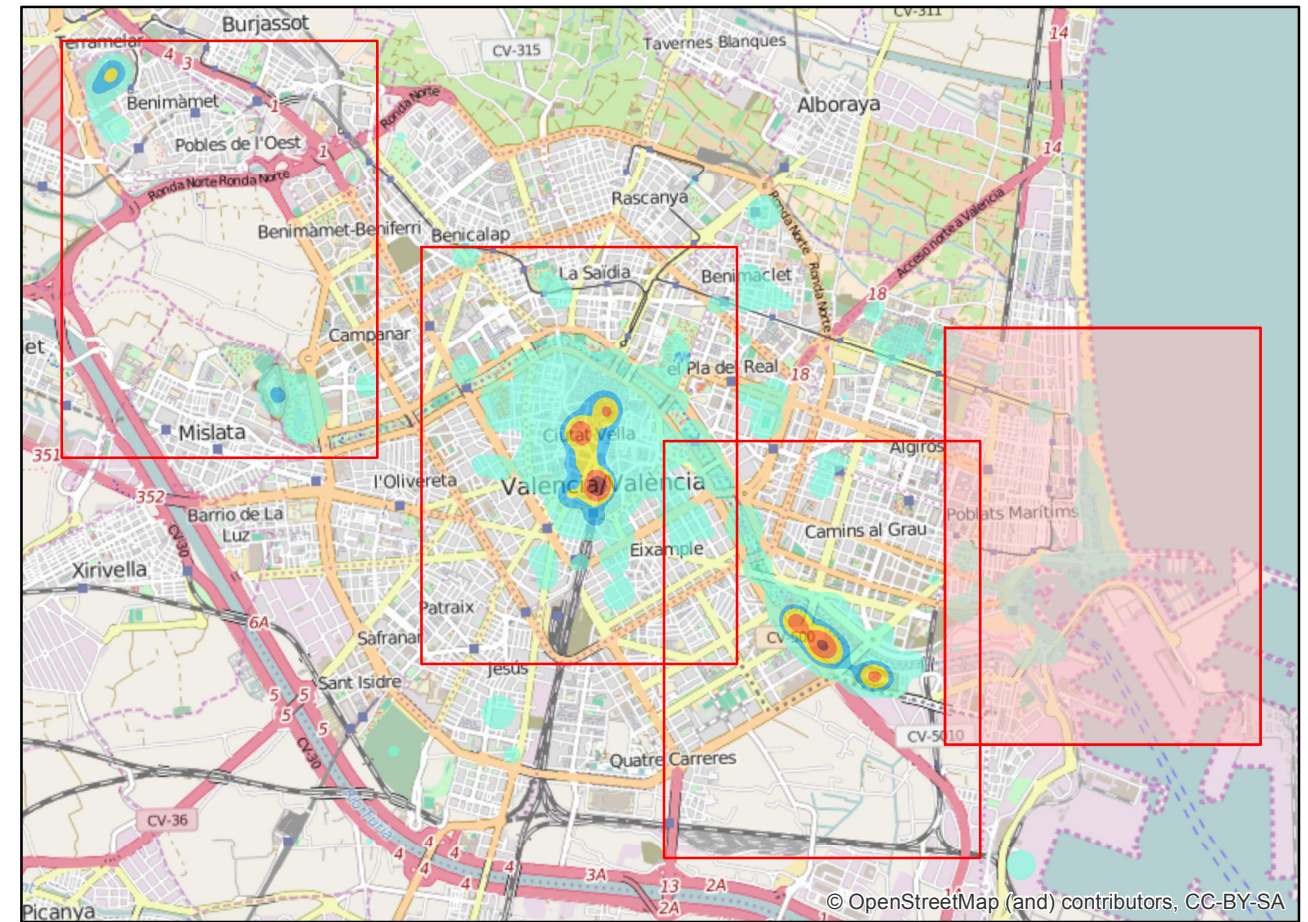
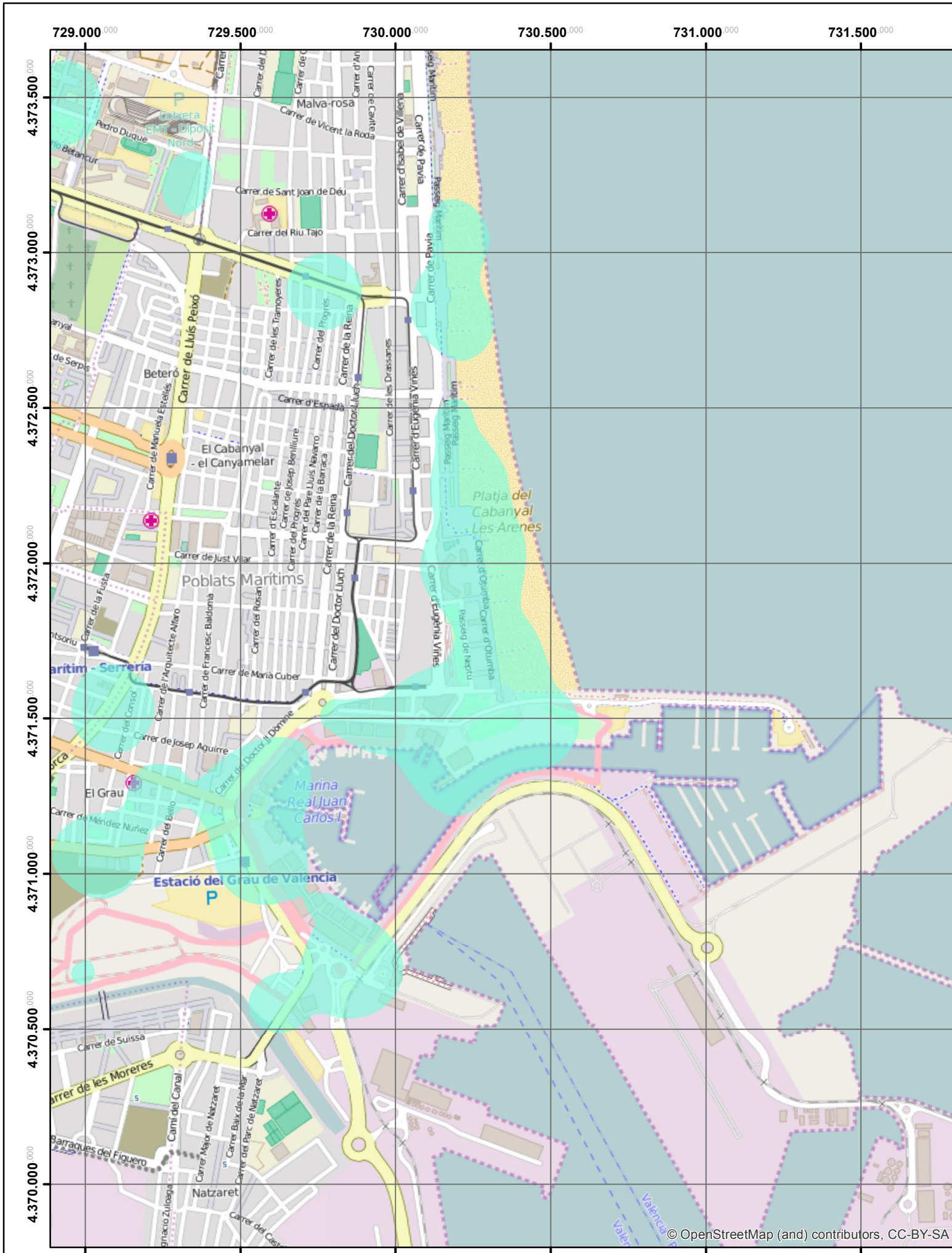
Nº Plano:

Autor: **Marcel Buenaventura Fenech**

Fecha: **29/08/2016**

2 / 4

**3. Plano de densidad de fotografías/videos de usuarios en
valencia. Puerto deportivo y playa de la Malvarrosa.**



Plano general

LEYENDA

- Hojas
- Hoja seleccionada

Mapa de densidad
Fotografías por m²

- No data
- 0,001 - 0,01
- 0,01 - 0,02
- 0,02 - 0,035
- 0,035 - 0,06
- 0,06 - 0,09

Características de los planos

- Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)
- Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte
- EPSG (European Petroleum Survey Group) : 25830
- Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



Título: **Plano de densidad de fotografías/videos de usuario en Valencia. Puerto deportivo y playa de la Malvarrosa.**

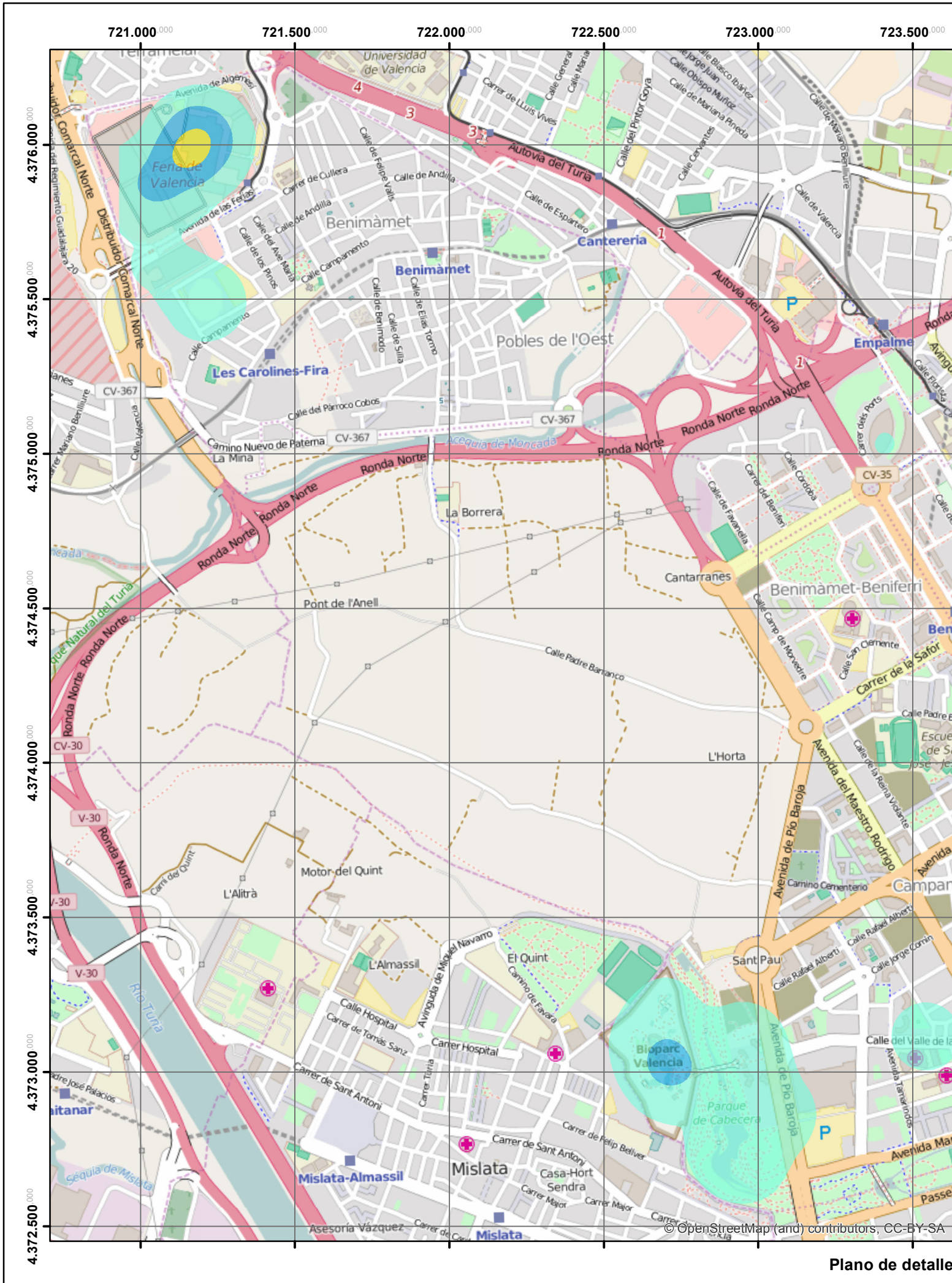
Nº Plano: **3 / 4**

Autor: **Marcel Buenaventura Fenech**

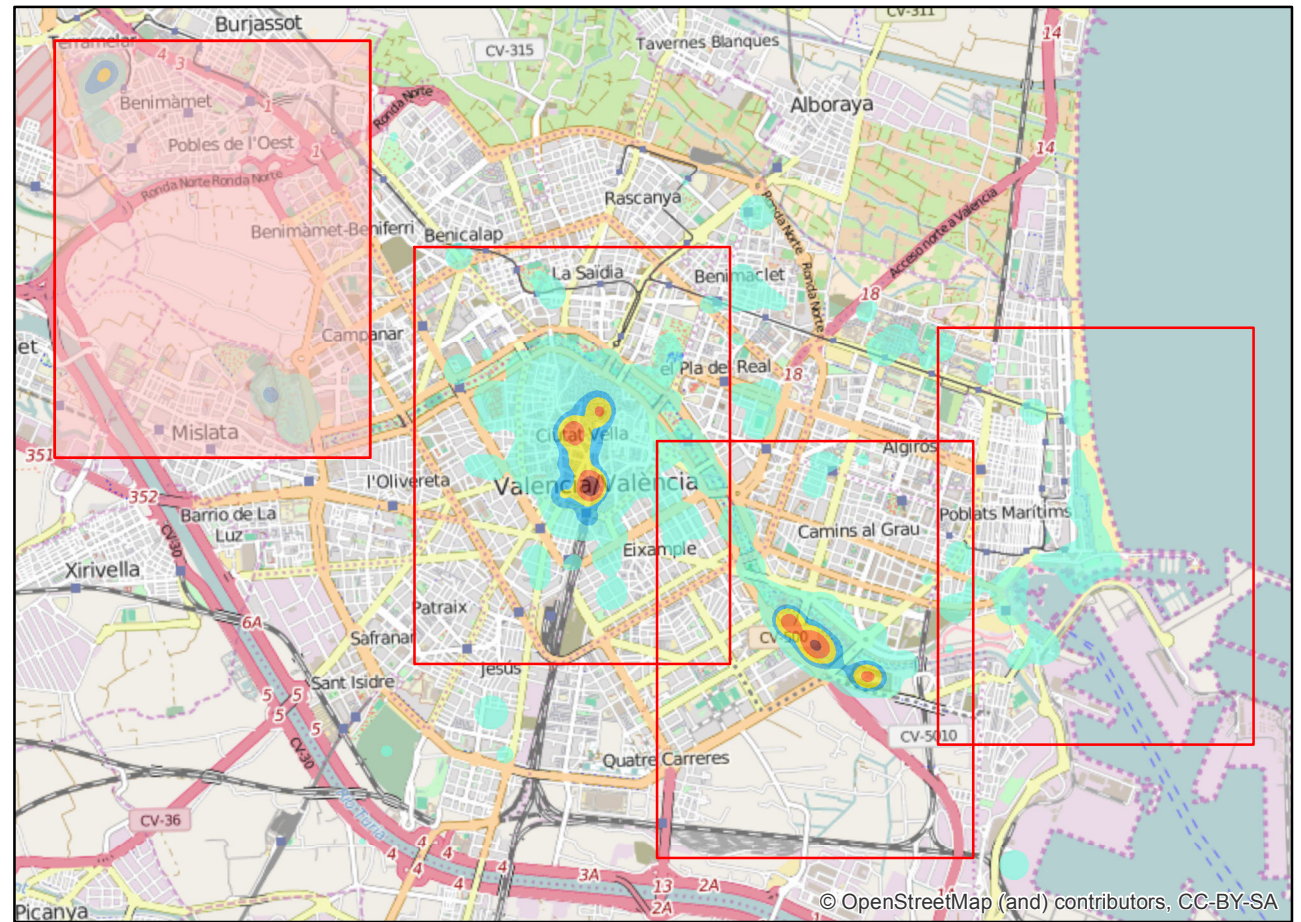
Fecha: **29/08/2016**

Plano de detalle

**4. Plano de densidad de fotografías/videos de usuarios en
valencia. Parque de Cabecera y Feria de Valencia
(Benimámet).**



Plano de detalle



Plano general

LEYENDA

- Hojas
- Hoja seleccionada

Mapa de densidad

Fotografías por m²

- No data
- 0,001 - 0,01
- 0,01 - 0,02
- 0,02 - 0,035
- 0,035 - 0,06
- 0,06 - 0,09

Características de los planos

- Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)
- Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte
- EPSG (European Petroleum Survey Group) : 25830
- Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



Título: **Plano de densidad de fotografías/videos de usuario en Valencia. Parque de Cabecera y Feria de Valencia (Benimàmet).**

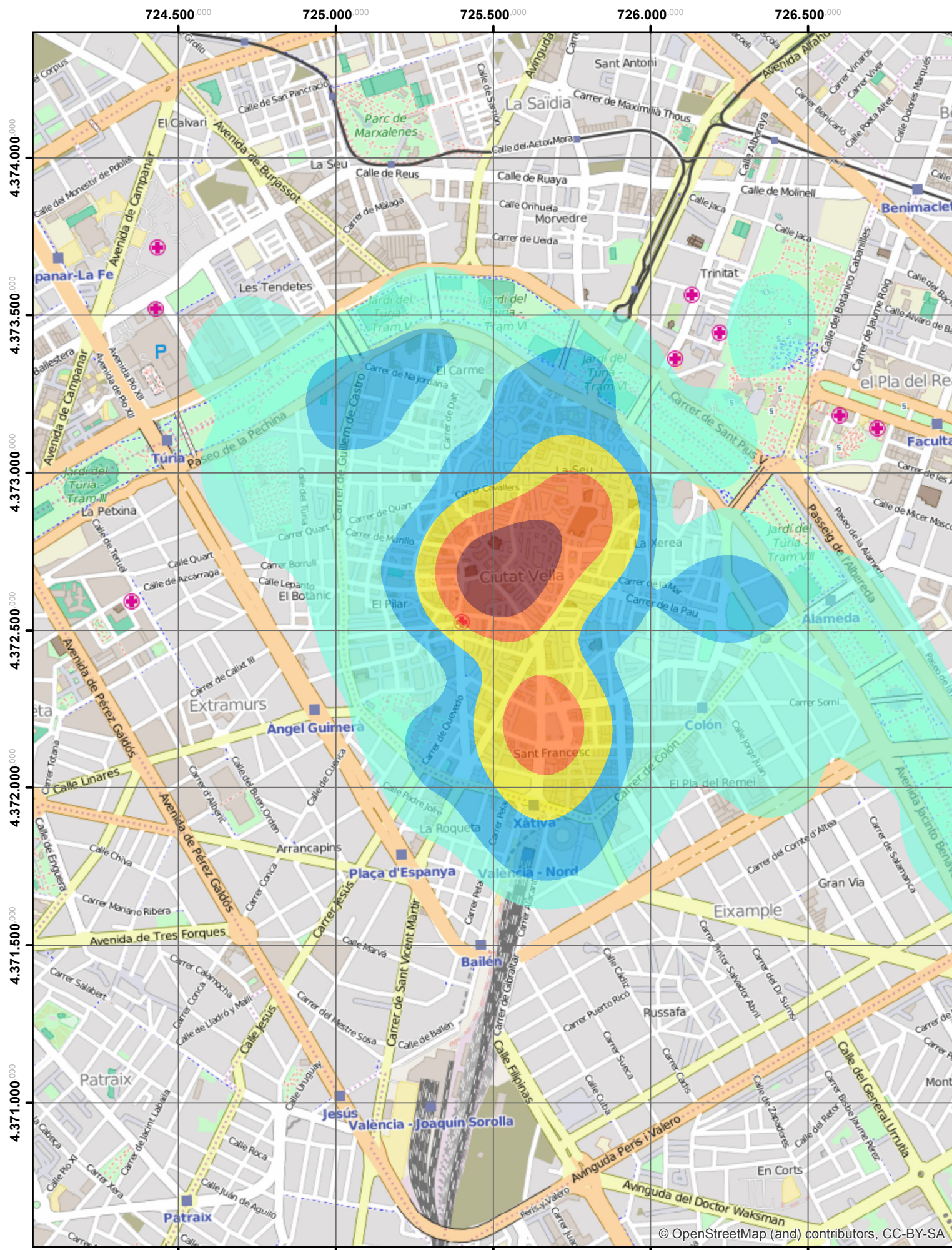
Nº Plano: **4 / 4**

Autor: **Marcel Buenaventura Fenech**

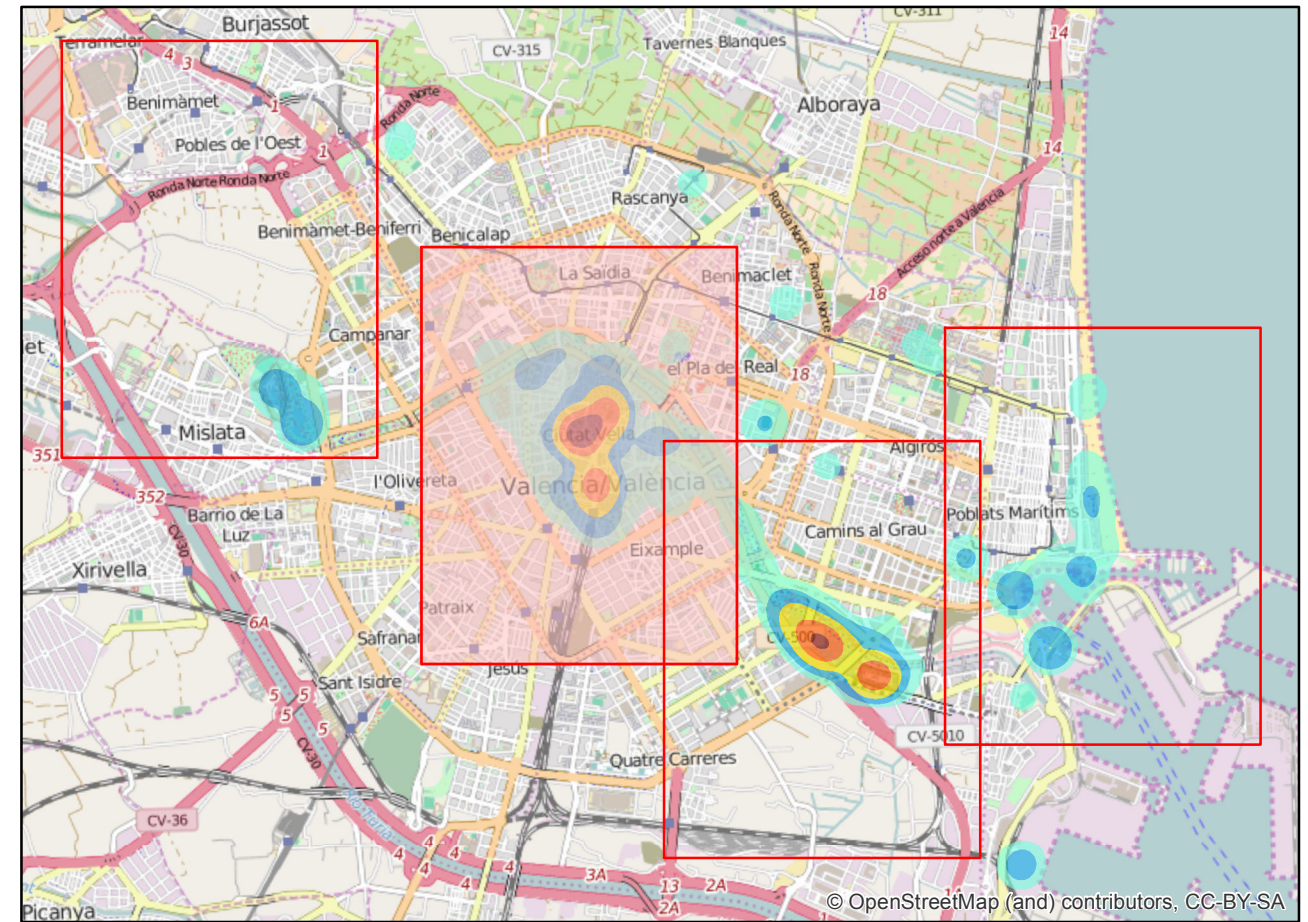
Fecha: **29/08/2016**

ANEXO II - Planos de turistas

1. Plano de densidad de fotografías/videos de turistas en Valencia. Casco histórico y alrededores.



Plano de detalle



Plano general

LEYENDA

- Hojas
- Hoja seleccionada

Mapa de densidad

Fotografías por m²

- No data
- 0,0003 - 0,001
- 0,001 - 0,003
- 0,003 - 0,007
- 0,007 - 0,012
- 0,012 - 0,02

Características de los planos

- Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)
- Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte
- EPSG (European Petroleum Survey Group) : 25830
- Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000

UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial

ESCUELA TÉCNICA SUPERIOR DE INGENIERÍA GEODÉSICA CARTOGRÁFICA Y TOPOGRÁFICA

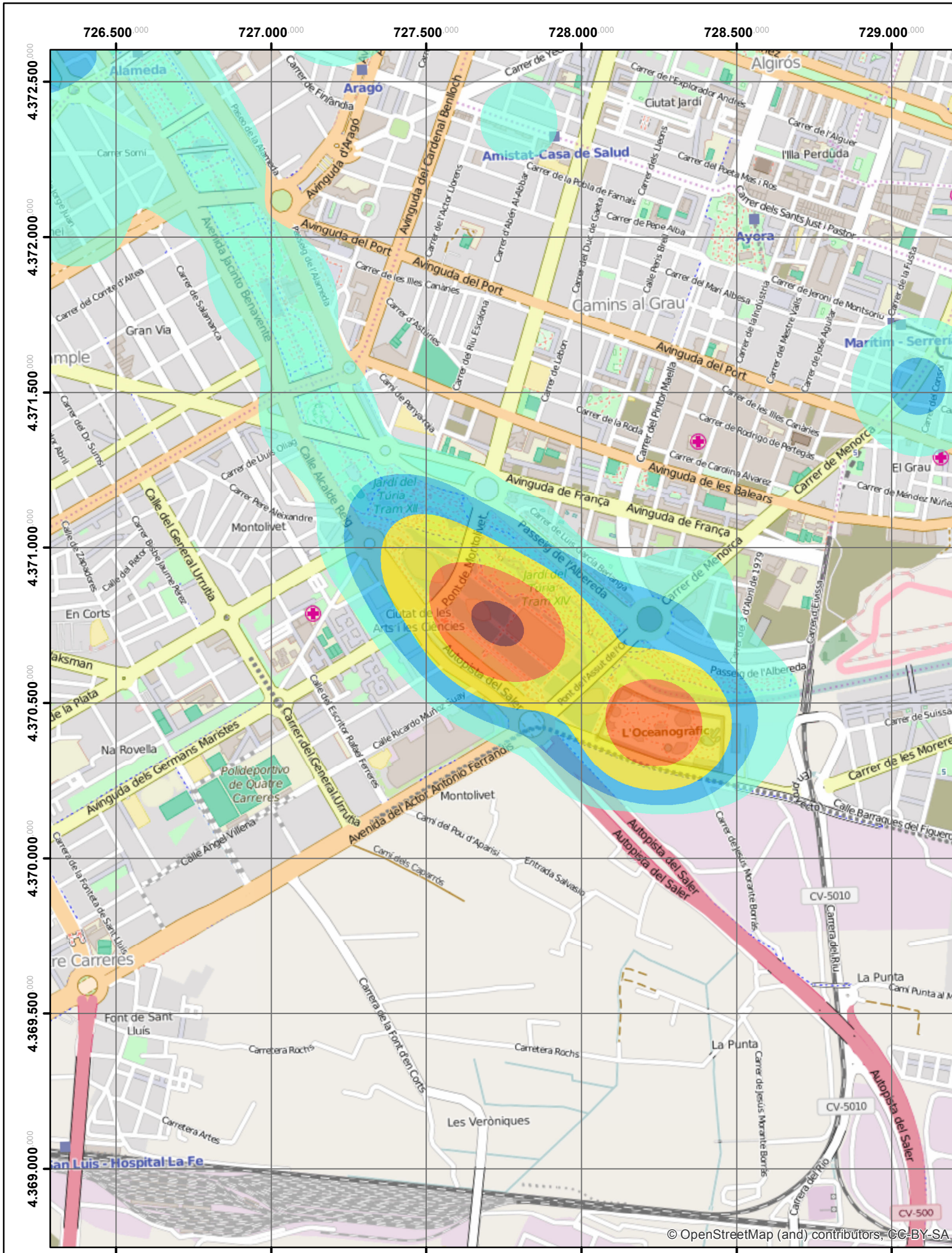
Título: **Plano de densidad de fotografías/videos de turista en Valencia. Casco histórico y alrededores.**

Autor: **Marcel Buenaventura Fenech**

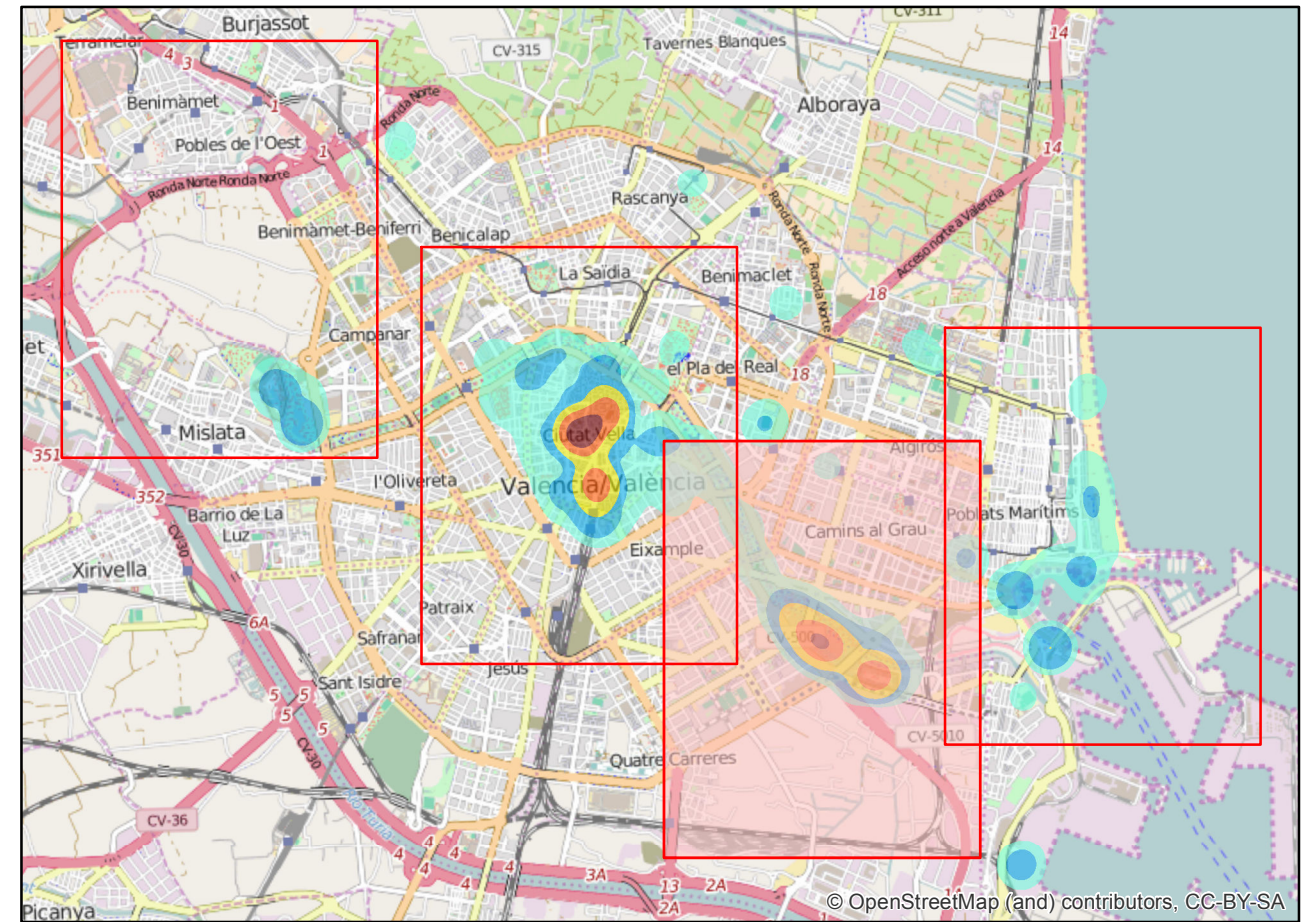
Nº Plano: **1 / 4**

Fecha: **29/08/2016**

2. Plano de densidad de fotografías/videos de turistas en Valencia. Ciudad de las Artes y las Ciencias.



Plano de detalle



Plano general

LEYENDA

- Hojas
- Hoja seleccionada
- Mapa de densidad**
- Fotografías por m²**
- No data
- 0,0003 - 0,001
- 0,001 - 0,003
- 0,003 - 0,007
- 0,007 - 0,012
- 0,012 - 0,02

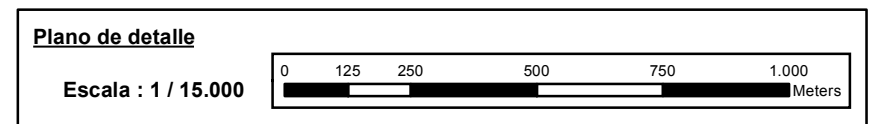
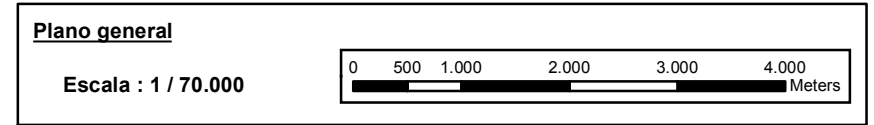
Características de los planos

Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)

Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte

EPSG (European Petroleum Survey Group) : 25830

Mapa base : OSM (Open Street Maps)



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



Título: **Plano de densidad de fotografías/videos de turista en Valencia. Ciudad de las Artes y las Ciencias.**

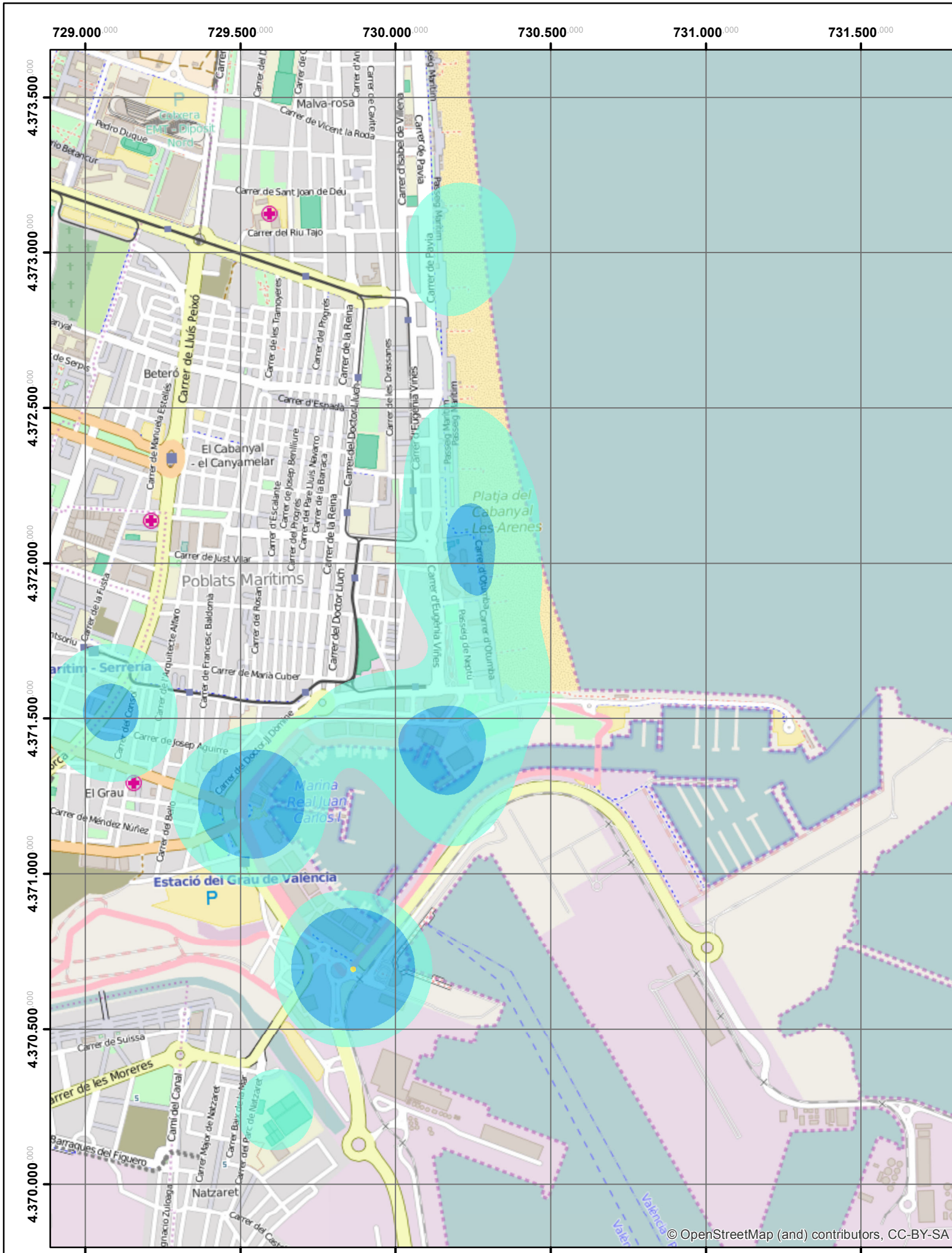
Nº Plano: **2 / 4**

Autor: **Marcel Buenaventura Fenech**

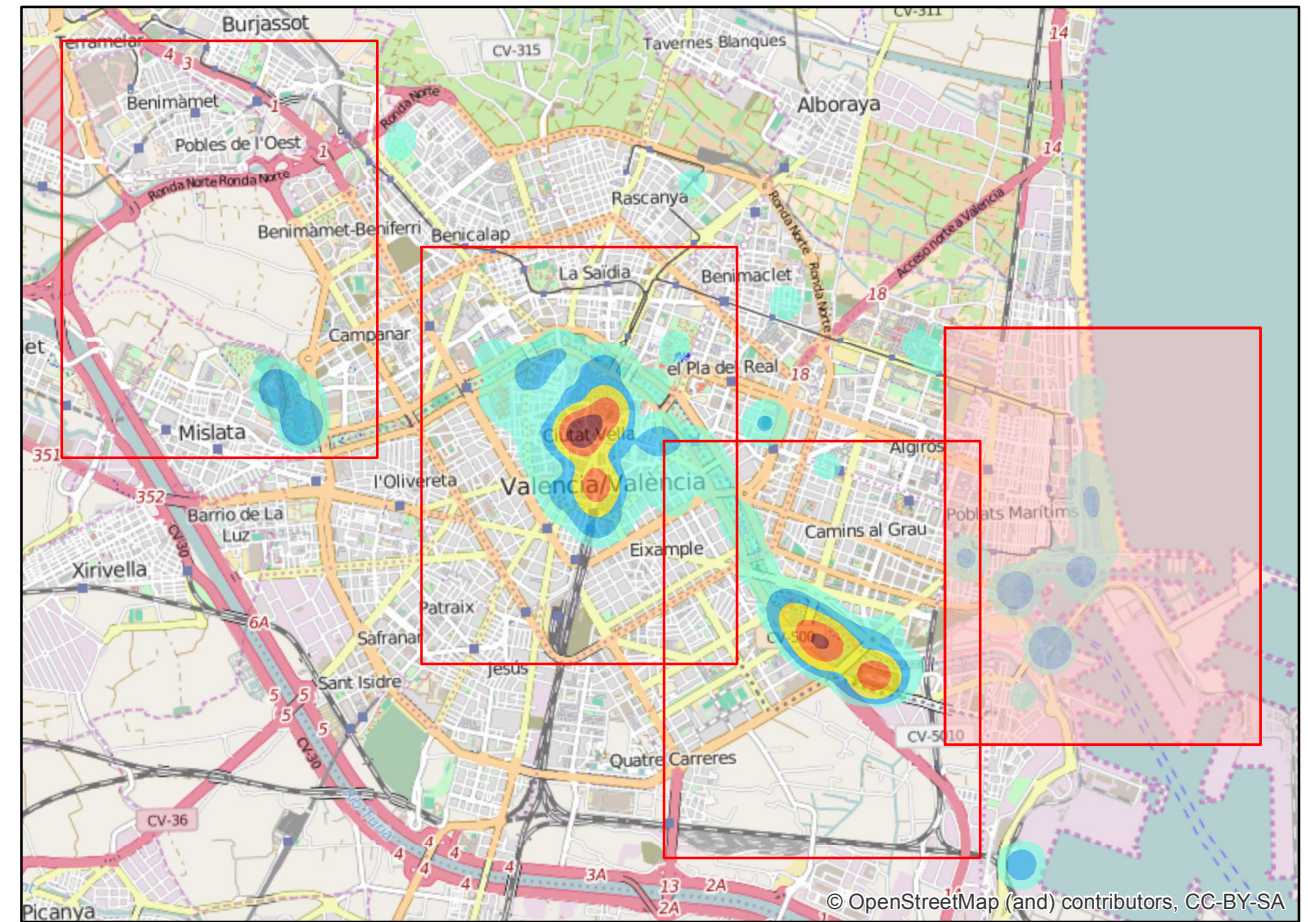
Fecha: **29/08/2016**

© OpenStreetMap (and) contributors, CC-BY-SA

3. Plano de densidad de fotografías/videos de turistas en valencia. Puerto deportivo y playa de la Malvarrosa.



Plano de detalle



Plano general

LEYENDA

- Hojas
 - Hoja seleccionada
- Mapa de densidad**
- Fotografías por m²**
- No data
 - 0,0003 - 0,001
 - 0,001 - 0,003
 - 0,003 - 0,007
 - 0,007 - 0,012
 - 0,012 - 0,02

Características de los planos

Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)

Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte

EPSG (European Petroleum Survey Group) : 25830

Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



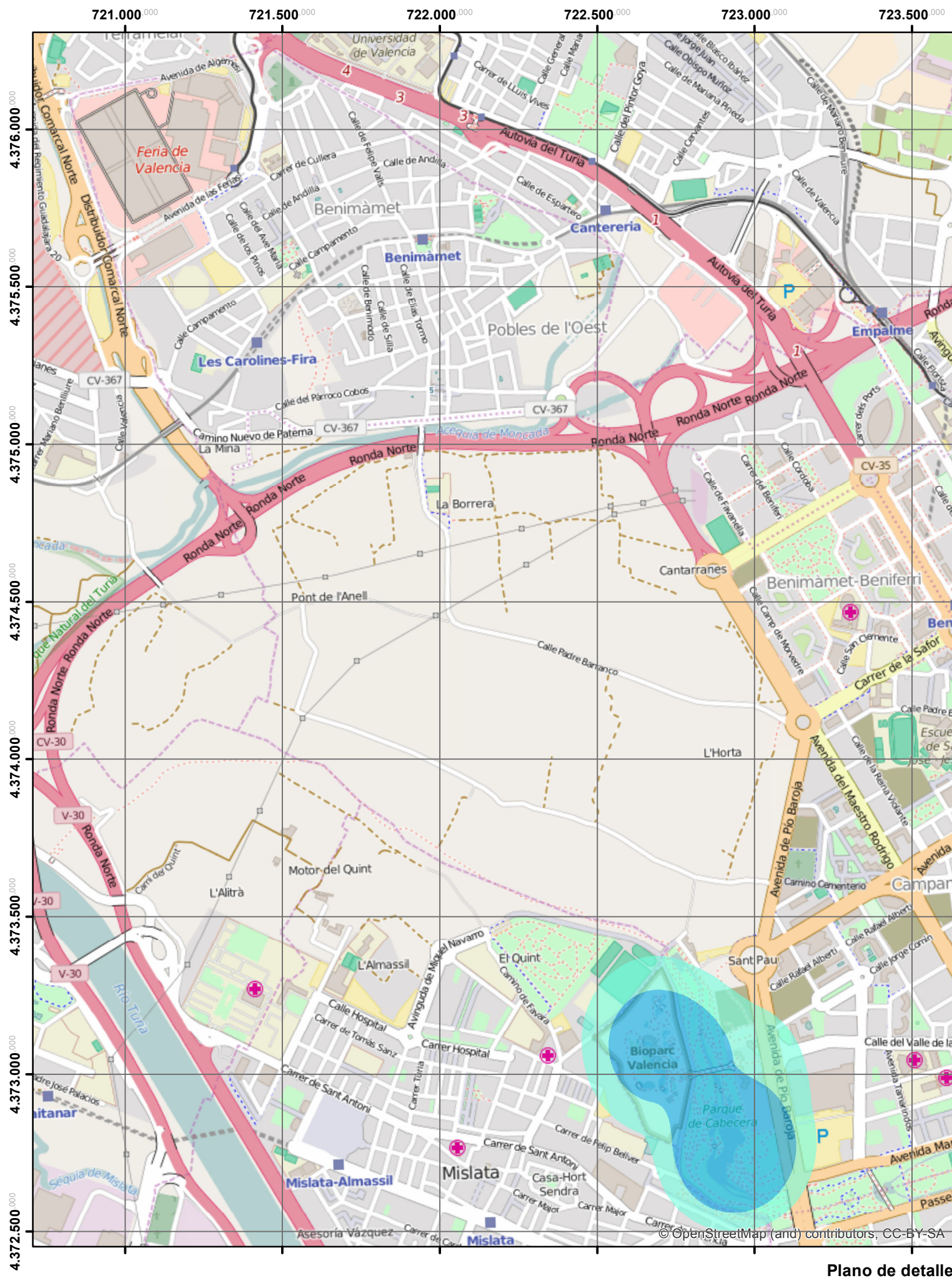
Título: **Plano de densidad de fotografías/videos de turista en Valencia. Puerto deportivo y playa de la Malvarrosa.**

Nº Plano: **3 / 4**

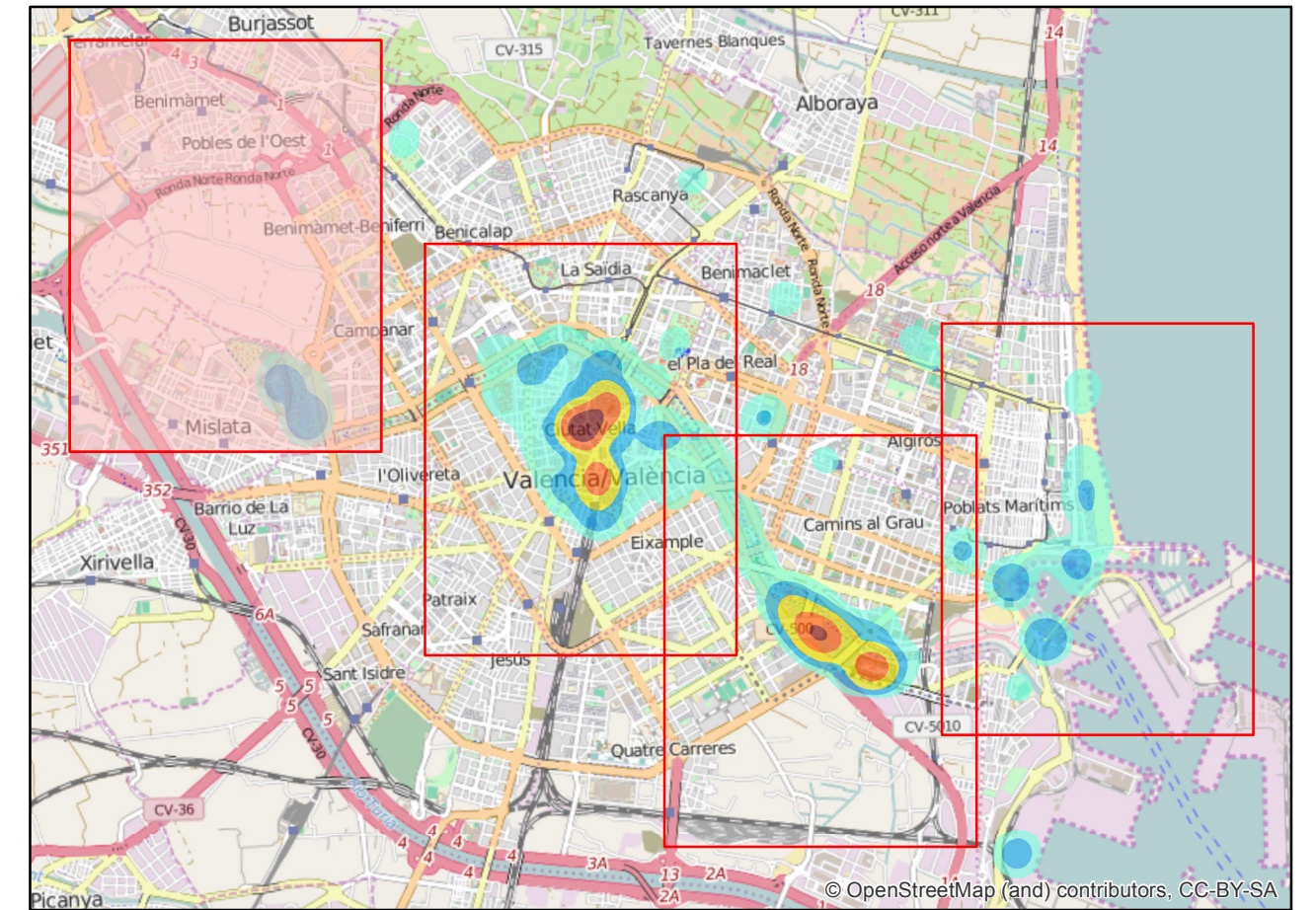
Autor: **Marcel Buenaventura Fenech**

Fecha: **29/08/2016**

**4. Plano de densidad de fotografías/videos de turistas en
valencia. Parque de Cabecera y Feria de Valencia
(Benimámet).**



Plano de detalle



Plano general

LEYENDA

- Hojas
- Hoja seleccionada

Mapa de densidad

Fotografías por m²

- No data
- 0,0003 - 0,001
- 0,001 - 0,003
- 0,003 - 0,007
- 0,007 - 0,012
- 0,012 - 0,02

Características de los planos

- Sistema de referencia : ETRS89 (European Terrestrial Reference System 1989)
- Proyección : UTM (Universal Transverse Mercator) Zona 30 Norte
- EPSG (European Petroleum Survey Group) : 25830
- Mapa base : OSM (Open Street Maps)

Plano general

Escala : 1 / 70.000

Plano de detalle

Escala : 1 / 15.000



Análisis del comportamiento de los usuarios de la red Flickr desde el punto de vista geosocial



Título: **Plano de densidad de fotografías/videos de turista en Valencia. Parque de Cabecera y Feria de Valencia (Benimàmet).**

Nº Plano: **4 / 4**

Autor: **Marcel Buenaventura Fenech**

Fecha: **29/08/2016**

