



Secretaría de Estado de Infraestructuras, Transporte y Vivienda

# Estudio de movilidad de viajeros de ámbito nacional aplicando la tecnología Big Data

## Informe metodológico

Versión 2 - 26 de julio de 2023

# Índice

<b>1. INTRODUCCIÓN .....</b>	<b>3</b>
1.1 ANTECEDENTES.....	3
1.2 OBJETO Y ALCANCE DEL ESTUDIO .....	4
1.3 GLOSARIO DE TÉRMINOS .....	6
<b>2. FUENTES DE DATOS UTILIZADAS .....</b>	<b>7</b>
2.1 REGISTROS ANONIMIZADOS DE TELEFONÍA MÓVIL .....	7
2.2 USOS DEL SUELO .....	7
2.3 DATOS DE POBLACIÓN .....	8
2.4 DATOS DE LA RED DE TRANSPORTE .....	8
<b>3. SOLUCIÓN TÉCNICA Y METODOLOGÍA .....</b>	<b>9</b>
3.1 EXTRACCIÓN DE LOS REGISTROS DE TELEFONÍA MÓVIL .....	9
3.2 GENERACIÓN DE LOS INDICADORES DE MOVILIDAD .....	9
3.3 FIABILIDAD DE LOS RESULTADOS Y ERROR MUESTRAL .....	11
<b>4. ENTREGABLES.....</b>	<b>12</b>
<b>ANEXO I. CONSIDERACIONES METODOLÓGICAS CON RESPECTO AL ANÁLISIS DE MOVILIDAD DURANTE EL ESTADO DE ALARMA.....</b>	<b>15</b>

## Registro de revisiones

Versión	Fecha	Modificación	Secciones afectadas
1	20/12/2021	Versión inicial	
2	26/07/2023	El campo 'zona_residencia' del indicador de pernoctaciones pasa a ser el distrito censal de residencia, según la codificación INE	4

# 1. Introducción

## 1.1 Antecedentes

En su labor de mejorar la eficiencia en la planificación y la gestión de las infraestructuras y servicios de transporte, el Ministerio de Transportes, Movilidad y Agenda Urbana (en adelante MITMA), a través de la Secretaría de Estado de Infraestructuras, Transporte y Vivienda, tiene entre sus funciones la realización de los llamados ‘Estudios de prospectiva de demanda’. Estos estudios se consideran esenciales para evaluar el impacto de las actuaciones de infraestructuras, servicios o normativas sobre la evolución del sistema de transportes, y constituyen una información de gran valor para distintos tipos de usuarios, tanto públicos como privados. Uno de estos estudios es el **análisis de la movilidad interna de viajeros**.

En el pasado, el estudio de la movilidad de viajeros en España se realizaba a través de las encuestas Movilia, las cuales suponían un importante esfuerzo económico, técnico y humano. La última encuesta realizada con ese objeto, Movilia 2007, data de hace ya más de una década. En los últimos años, la proliferación de nuevas fuentes de datos geolocalizados procedentes de dispositivos móviles personales ha dado lugar a nuevas formas de estudiar la movilidad de las personas y obtener información de demanda de transporte en un plazo de tiempo reducido y a un coste significativamente menor que el de los métodos tradicionales, eliminando o mitigando algunas de las principales limitaciones de las encuestas de movilidad. Los datos procedentes de la red de telefonía móvil resultan particularmente interesantes para este propósito, gracias a la posibilidad de obtener muestras de gran tamaño y con una elevada resolución espacio-temporal de prácticamente todos los segmentos de población. Con el fin de aprovechar estas oportunidades, la Secretaría de Estado de Infraestructuras, Transporte y Vivienda lanzó en 2017 el **‘Estudio de la movilidad interprovincial de viajeros aplicando la tecnología big data’**<sup>1</sup>, un proyecto pionero que empleó los registros generados por los terminales móviles al conectarse a las redes de telefonía como fuente principal de datos para estudiar la movilidad de viajeros a escala nacional. Aprovechando la experiencia adquirida en el estudio anterior, en 2020 el MITMA llevó a cabo el estudio **‘Análisis de la movilidad en España con tecnología big data durante el Estado de Alarma para la gestión de la crisis del COVID-19’**<sup>2</sup>, en el que se emplearon datos de telefonía móvil para monitorizar la movilidad de manera diaria y apoyar distintos procesos de toma de decisiones relativas a la gestión de la emergencia sanitaria ocasionada por la pandemia de COVID-19.

Con el precedente de estos dos proyectos, el MITMA ha lanzado el presente proyecto denominado **‘Estudio de Movilidad de Viajeros de ámbito nacional aplicando la tecnología Big Data’** que dará continuidad a los anteriores, ampliando su alcance y profundizando en esta línea de trabajo.

---

<sup>1</sup> <https://observatoriotransporte.mitma.es/estudio-experimental>

<sup>2</sup> <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data>

## 1.2 Objeto y alcance del estudio

El presente documento recoge el alcance, metodología y entregables generados. Las especificaciones se detallan a continuación:

- **Viajes objeto de estudio:** viajes de más de 500 metros con origen y/o destino en el territorio nacional.
- **Población de estudio:** residentes en España.
- **Período de estudio:** la duración del periodo de estudio es de un mínimo de 18 meses, que podrá prolongarse hasta un máximo de otros 18 meses adicionales. Se analiza la movilidad desde el día 1 de enero de 2022 en adelante.
- **Zonificación:** la zonificación base de estudio se compone de un conjunto de zonas que constituyen agregaciones de distritos censales para el caso del territorio de España y zonas NUTS-3 para Francia y Portugal. Esta zonificación presenta un total de 3.909 zonas. Para los viajes con origen o destino fuera de la zonificación de estudio se ha definido una zona externa que cubre el resto del mundo. Partiendo de esta zonificación base, los resultados también se presentan agregados en dos zonificaciones adicionales para facilitar la explotación de la información: una a nivel municipal con un total de 2.735 zonas y otra a nivel de grandes áreas urbanas (GAUs) con un total de 2.203 zonas.
- **Resolución temporal:** los viajes se segmentan por franjas horarias de 1 hora de duración. Como criterio de asignación de la hora del viaje se considera la hora de inicio del mismo en el caso de viajes desde España. Para los viajes con origen fuera del territorio nacional se considera como hora del viaje la hora de entrada en España.
- **Distancia de viaje:** se determina como la distancia ortodrómica entre el origen y el destino del viaje y se emplean los siguientes intervalos: 500 metros-2 km, 2-10 km, 10-50 km y >50 km. Para el caso de viajes al/desde el extranjero, la distancia corresponde a la distancia entre el origen/destino del viaje y el punto de salida/entrada en el territorio nacional.
- **Segmentación por propósito de viaje:** para cada viaje, se determina el tipo de actividad en el origen y el destino, distinguiendo entre casa, trabajo/estudio, otras actividades frecuentes, y actividades esporádicas. Adicionalmente, se incorpora un indicador adicional para señalar si la actividad potencialmente pudiera estar relacionada con propósito estudio. Es importante señalar que la categoría 'trabajo/estudio' contiene principalmente trabajos/estudios a tiempo completo en ubicaciones fijas. Parte de la movilidad obligada puede también estar recogida en el grupo 'otras actividades frecuentes', para aquellos casos de actividades laborales o estudiantiles que no hayan sido clasificadas en el grupo anterior al tener un patrón también compatible con otro tipo de actividades frecuentes (p.ej. actividades de ocio frecuentes).
- **Segmentación por perfil sociodemográfico<sup>3</sup>:** se segmentan los viajes de acuerdo a las siguientes características de los viajeros:

---

<sup>3</sup> Algunas de las variables sociodemográficas puede que no se proporcionen en algunos casos por criterios de privacidad; presentando el valor 'NA' en tal caso.

- Lugar de residencia a nivel provincial
- Edad, según las siguientes franjas etarias: 0-24, 25-44, 45-64 y  $\geq 65$ .
- Sexo, diferenciando entre 'hombre' y 'mujer'
- Nivel de renta, según las siguientes franjas en miles de euros:  $< 10$ , entre 10 y 15 y  $>15$ .

Para cada día de estudio se generan los siguientes productos principales:

- **Matriz de viajes básica**, que proporciona el número de viajes y viajes-km para cada combinación de par origen-destino, hora del día, intervalo de distancia de viaje, actividad en origen, actividad en destino y perfil sociodemográfico de los viajeros.
- **Viajes por persona**, que proporciona el número de personas que realizan un número determinado de viajes al día por zona de pernoctación, segmentando por rangos de edad y sexo.
- **Pernoctaciones**, que permiten obtener el número de personas que pernoctan en cada zona identificando su lugar de residencia

En la sección 4 de este documento se presenta una descripción detallada de cada uno de los productos/ficheros generados.

### 1.3 Glosario de términos

<b>Actividad</b>	Estancia de un individuo en un determinado lugar originada por una interacción o conjunto de interacciones con el entorno (incluyendo interacciones con otras personas, con un determinado servicio, etc.) que motivan que el individuo se desplace hasta allí.
<b>Estancia</b>	Permanencia de un individuo durante cierto tiempo en un lugar determinado.
<b>Etapas</b>	Desplazamiento entre las localizaciones de dos estancias consecutivas realizado en un modo de transporte. La utilización de modos de transporte colectivo lleva aparejada, por lo general, una etapa de acceso/egreso a/desde la terminal o la parada correspondiente. Dentro de un desplazamiento en un mismo modo de transporte pueden producirse transbordos que se considere que no dan lugar a dos etapas distintas: así, por ejemplo, es razonable considerar un desplazamiento en metro con un transbordo como una única etapa. Por otro lado, en ocasiones puede resultar útil identificar distintos trayectos consecutivos en un mismo modo de transporte; esta distinción es particularmente relevante para la validación y/o el ajuste de los resultados obtenidos con datos de telefonía mediante la comparación con datos de billeteo, en el caso de que los dos trayectos se realicen mediante servicios distintos y se disponga del billeteo de cada uno por separado.
<b>Tour</b>	Secuencia de actividades y viajes dentro de un mismo día que empieza y finaliza en una misma localización (habitualmente el hogar). Un tour está compuesto por dos o más viajes (por ejemplo, una secuencia de viajes hogar-trabajo-compras-hogar).
<b>Viaje</b>	Desplazamiento entre las localizaciones de dos actividades consecutivas. Un viaje tiene un propósito principal determinado por la actividad en el origen, la actividad en el destino o la combinación de ambas. Un viaje puede ser monomodal o multimodal y estar compuesto por una o varias etapas.

## 2. Fuentes de datos utilizadas

### 2.1 Registros anonimizados de telefonía móvil

La principal fuente de datos la constituyen los registros anonimizados de telefonía móvil. El estudio parte de una muestra de datos de más de 13 millones de líneas móviles proporcionada por Orange España.

Los datos de partida pueden clasificarse en tres categorías:

- **Datos de eventos registrados<sup>4</sup>:** datos anonimizados asociados a los registros de conexión de los dispositivos móviles con la red de telefonía móvil. Estos registros incluyen tanto eventos pasivos como activos. Los eventos pasivos están constituidos por lo que se denominan **datos de sondas de red**, que proporcionan una granularidad temporal muy elevada, generando un registro en España cada vez que el dispositivo cambia de antena. A estos registros se les unen los eventos activos denominados **CDRs** (Call Detail Records), que proporcionan un registro cada vez que el dispositivo interactúa con las redes móviles (llamadas, envío de mensajes de texto, sesiones de datos). En cuanto a la resolución espacial, se dispone de información de localización a nivel de celda de telefonía, lo que supone una precisión espacial de decenas o cientos de metros en ciudad y hasta varios kilómetros en zonas rurales.
- **Datos de la topología de la red de telefonía móvil:** datos sobre la red de telefonía, incluyendo la localización de las torres de comunicación y la orientación de las antenas.
- **Datos sociodemográficos:** datos sobre el sexo y rangos de edad de los titulares de las líneas móviles.

*Es importante señalar en este punto que los datos empleados en este proyecto son sustancialmente distintos a los datos utilizados en estudios anteriores, presentando estos datos una mejora de calidad en lo referente a granularidad temporal. En el Anexo I se comentan los aspectos a tener en consideración a la hora de comparar los resultados de este proyecto con estudios anteriores, señalando la influencia del cambio de fuente de datos para este proyecto.*

### 2.2 Usos del suelo

Se han utilizado también datos de usos del suelo, para mejorar la caracterización y la localización espacial de las actividades identificadas a partir de los datos de telefonía móvil. Los datos de usos del suelo proceden del Sistema de Información sobre Ocupación del Suelo de España (SIOSE) y de otras bases de datos disponibles a nivel autonómico. Adicionalmente, se han utilizado datos de ubicación de centros educativos procedentes de los POIS (Puntos de interés) del producto BTN (Base Topográfica Nacional) del Instituto Geográfico Nacional para la caracterización de las posibles actividades asociadas a estudio.

<sup>4</sup> Los datos de eventos registrados se procesan en un entorno seguro en la infraestructura del operador móvil para generar información agregada y por tanto anonimizada, con el fin de cumplir con lo establecido en la LOPD-GDD.



## 2.3 Datos de población

Para los procesos de elevación de la muestra se han utilizado datos procedentes del Padrón Municipal de Habitantes.

## 2.4 Datos de la red de transporte

Los algoritmos empleados para la identificación de viajes emplean también información de la red de transporte (por ejemplo, localización de aeropuertos, red ferroviaria, etc.), con el objetivo de refinar la distinción entre actividades (que puedan suponer el inicio o fin de un viaje) y paradas intermedias (que no siendo fin de viaje puedan diferenciar las etapas del mismo).

## 3. Solución técnica y metodología

### 3.1 Extracción de los registros de telefonía móvil

El primer subproceso consiste en la extracción y pseudonimización de los registros de telefonía móvil. La pseudonimización de los registros está basada en la utilización de una función hash unidireccional, es decir, una función que permite el cálculo de un identificador anonimizado (similar a un texto aleatorio) a partir del identificador original (habitualmente el IMSI, en el caso de un operador de telefonía) de tal forma que resulta imposible realizar el proceso a la inversa. Se utiliza lo que se conoce como funciones hash perfectas, que por su diseño evitan las colisiones, es decir, evitan que dos identificadores originales diferentes den como resultado un mismo identificador anonimizado. Una vez anonimizados, los registros de telefonía se almacenan en un entorno seguro dentro de la infraestructura del operador móvil, en el que se instala el software necesario para generar los indicadores agregados y anonimizados.

### 3.2 Generación de los indicadores de movilidad

La generación de los indicadores de movilidad se ha llevado a cabo utilizando un software especializado desarrollado a tal efecto. Este software ha sido empleado en más de 150 proyectos en distintos países en los que se han utilizado datos anonimizados de telefonía móvil para la caracterización de la movilidad urbana e interurbana, tanto para clientes públicos (agencias estadísticas, autoridades de transporte, etc.) como privados (empresas concesionarias de autopistas, operadores de autobuses interurbanos, consultoras de transporte, etc.). Entre estos proyectos se incluye los ya mencionados 'Estudio de la Movilidad Interprovincial de Viajeros aplicando la Tecnología Big Data' llevado a cabo por el Ministerio de Transportes, Movilidad y Agenda Urbana en el año 2018 y el 'Análisis de la movilidad en España con tecnología big data durante el Estado de Alarma para la gestión de la crisis del COVID 19'

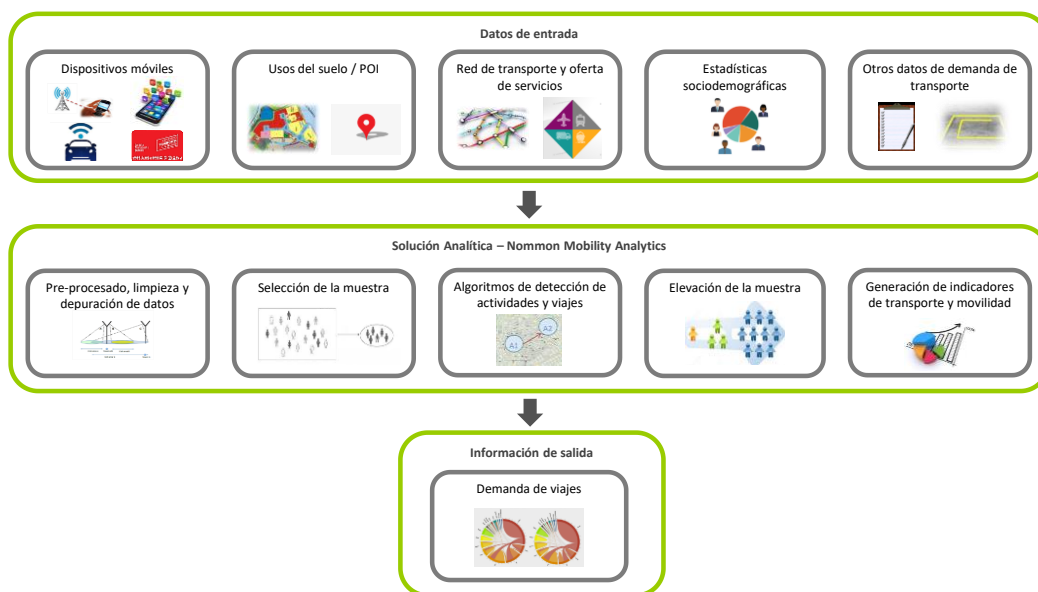


Figura 1 – Esquema de alto nivel de la solución técnica empleada en el proyecto

La Figura 1 muestra un esquema de alto nivel de la solución técnica. El procesado y análisis de los datos consta de los subprocesos principales descritos a continuación. Como ya se ha señalado anteriormente, todos estos procesos se llevan a cabo dentro de la infraestructura del operador móvil, de manera que la información generada y entregada al Ministerio es ya información agregada y anonimizada.

1. **Pre-procesado y limpieza de los datos.** En primer lugar se realiza un pre-procesado de los datos de telefonía para facilitar su gestión, ordenando y agrupando los registros de la forma más conveniente para su posterior análisis. También se lleva a cabo un análisis de integridad de los datos para eliminar posibles errores en los datos del operador móvil. Este proceso resulta esencial para asegurar la calidad de los datos, evitando que posibles errores de origen desvirtúen los resultados que se obtienen con los algoritmos de extracción de patrones de actividad y movilidad.
2. **Construcción de la muestra.** Para construir la muestra se realiza una selección de los usuarios válidos para proporcionar información relativa a sus desplazamientos. Dicha selección se realiza de acuerdo con distintos criterios relacionados con su actividad telefónica, de manera que ésta sea suficiente para establecer sus patrones de comportamiento con un nivel de fiabilidad adecuado. La construcción de la muestra supone un compromiso entre cantidad y calidad. Ejercicios de validación llevados a cabo en proyectos anteriores demuestran la importancia de seleccionar una muestra de buena calidad, aún a costa de reducir ligeramente el tamaño muestral, para evitar la inclusión de usuarios que realicen actividades y viajes imposibles de detectar y que puedan por tanto afectar a la calidad de las matrices origen-destino y del resto de indicadores a generar.
3. **Identificación de actividades frecuentes y el lugar de pernoctación.** A partir del análisis de los hábitos de comportamiento de los usuarios a lo largo de varias semanas se identifican las actividades frecuentes realizadas por el usuario, como su lugar de residencia habitual (el cual se utilizará posteriormente en el proceso de elevación muestral), su lugar de trabajo/estudio, otras actividades frecuentes o actividades esporádicas. Adicionalmente, con el objetivo de diferenciar entre actividades laborales y las asociadas a estudio, se han analizado patrones asociados a la cercanía de la actividad a centros educativos, el horario de la actividad y el perfil socio-demográfico del usuario con el objetivo de valorar la potencialidad de que la actividad sea de carácter educativo. Por último, también se ha identificado la actividad asociada al lugar de pernoctación de los usuarios en el día de estudio.
4. **Extracción de actividades y viajes.** Para identificar actividades y viajes, se emplea una combinación de criterios basados en los tiempos de estancia, los itinerarios de los desplazamientos y los patrones de comportamiento a lo largo del periodo de estudio, filtrando las estancias intermedias subordinadas al viaje y realizadas entre etapas del mismo (por ejemplo, una parada intermedia para realizar un transbordo entre autobuses). El resultado de este proceso es la secuencia de actividades y viajes realizados por cada usuario en los días de estudio. La información asociada a cada actividad incluye su localización (a nivel de celda de telefonía móvil), el tipo de actividad (casa, trabajo/estudio, etc.), la hora de inicio de la actividad y la hora de finalización. La información asociada a cada viaje incluye origen (localización de la actividad inmediatamente anterior al viaje), destino (localización de la

actividad inmediatamente posterior al viaje), hora de inicio del viaje (hora de finalización de la actividad anterior) y hora de finalización (hora de inicio de la actividad siguiente).

5. **Elevación de la muestra.** La expansión de la muestra se realiza tomando como marco muestral la población residente en el país, según los datos del Padrón de Habitantes proporcionados por el INE. Se emplean procedimientos estándar de elevación muestral (similares a los que se emplean, por ejemplo, en una encuesta domiciliaria de movilidad), aplicando factores de expansión estratificados por lugar de residencia a nivel de distrito censal, edad y género. Además, se aplica un criterio mínimo de tamaño muestral, descartando aquellos estratos para los que la muestra es inferior al 1% de la población (es decir, para los que el factor de expansión es superior a 100), evitando así que factores de elevación excesivamente altos puedan distorsionar los indicadores de movilidad.
6. **Generación de indicadores.** Finalmente, la información obtenida se agrega con la resolución espacial y temporal requerida y las segmentaciones deseadas para generar las matrices origen-destino y el resto de indicadores de movilidad. La agregación se realiza de tal forma que el tamaño poblacional de los distintos grupos de población analizados garantice la imposibilidad de reidentificar a ningún individuo mediante un hipotético proceso de fusión con otras fuentes de datos, de acuerdo con los requisitos de la LOPD-GDD. Por otro lado, teniendo en cuenta el criterio de limitación de los factores de elevación muestral descrito en el punto 5, cuando para una determinada zona se ha descartado más del 20% del marco muestral, no se proporcionan los indicadores correspondientes a dicha zona. Adicionalmente, para cada producto, se pueden aplicar criterios de privacidad adicionales (ver detalle en 4.Entregables).

### 3.3 Fiabilidad de los resultados y error muestral

Se asume que la muestra de los usuarios de uno de los tres principales operadores en cada zona del territorio y para cada estrato sociodemográfico se aproxima razonablemente bien a una muestra aleatoria de la población residente en dicha zona, salvo por las limitaciones intrínsecas asociadas a la tecnología (ausencia de niños de muy corta edad, que no disponen de teléfono móvil, y menor representación de los ancianos de edad avanzada, algunos de los cuales tampoco son usuarios de líneas móviles). En estas condiciones, y en base a la experiencia de numerosos estudios de movilidad llevados a cabo en los últimos años por numerosas autoridades de transporte a nivel nacional, autonómico y municipal, se considera que la muestra utilizada, de más de 13 millones de líneas móviles, proporcionará un alto nivel de fiabilidad para los indicadores de movilidad a nivel de CCAA y provincia, así como para la movilidad de los municipios de mayor tamaño y las principales relaciones de movilidad entre municipios, suficiente para cumplir con los objetivos del estudio. El error muestral aumentará a medida que se toman resultados más desagregados (por ejemplo, movilidad en municipios pequeños), así como en las relaciones con menor número de viajes. Del mismo modo, es importante señalar que los resultados de movilidad a nivel nacional presentan mayor fiabilidad que los resultados de viajes con origen o destino el extranjero, motivado por la menor información disponible en estos casos (número mucho más reducidos (o inexistente para algunos usuarios) de eventos de red fuera del territorio nacional y una precisión espacial muy inferior).

## 4. Entregables

Los entregables se describen a continuación:

1. **Zonificaciones de estudio.** En el proyecto, aparte de la zonificación principal de estudio (3.743 zonas), se han definido un conjunto de agregaciones relevantes con el objetivo de facilitar la explotación de los resultados. En concreto, se han realizado agregaciones a nivel municipal (o agregaciones de municipios) y a nivel de grandes áreas urbanas (GAUs). Se han generado los siguientes ficheros de zonificación en formato digital:
  - a. **'zonificacion\_distritos.shp'**, zonificación de mayor detalle compuesta por distritos o agregaciones de distritos.
  - b. **'zonificacion\_municipios.shp'**, agregación de la zonificación de detalle a nivel municipal o agregaciones de municipios.
  - c. **'zonificacion\_GAU.shp'**, agregación de la zonificación de detalle a nivel de grandes áreas urbanas y municipios.

Asociados a estos ficheros se dispone de los ficheros de nombres de las zonas y de población:

- **'relacion\_ine\_zonificacionMitma.csv'**, fichero que relaciona la codificación del INE con las distintas codificaciones empleadas en las distintas zonificaciones del proyecto
  - **'poblacion.csv'**, un fichero con información de población por distrito censal y su relación con municipio y provincia según codificación INE.
  - **'poblacion\_[zona].csv'**, un fichero por cada zonificación ('distritos', 'municipios' y 'GAU') que proporciona para cada zona la población residente.
  - **'nombres\_[zona].csv'**, un fichero por cada zonificación ('distritos', 'municipios' y 'GAU') que proporciona para cada zona su nombre.
2. **Matrices de viajes.** Ficheros en formato 'CSV' con la información de movilidad diaria de los residentes en España. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es 'AAAAMMDD\_Viajes\_[zona].csv.gz', donde 'AAAAMMDD' es el año, mes y día de estudio y '[zona]' el tipo de zonificación empleada pudiendo tomar los valores 'distritos', 'municipios' y 'GAU'. La matriz de viajes presenta los siguientes campos:
    - a. **'fecha'**: fecha de estudio en formato AAAAMMDD
    - b. **'periodo'**: periodo horario en 24 franjas. '00' significa la movilidad con origen entre las 00:00:00 y las 00:59:59. Para los viajes con origen fuera del territorio nacional, la hora del viaje es la hora de entrada en España.
    - c. **'origen'**: identificador de la zona de origen. Para los viajes con origen fuera de la zonificación de estudio, se define el origen "externo".
    - d. **'destino'**: Identificador de la zona de destino. Para los viajes con destino fuera de la zonificación de estudio, se define el destino "externo".
    - e. **'distancia '**: distancia ortodrómica (km) entre el origen y el destino del viaje en rangos: 500 metros-2 km, 2-10 km, 10-50 km y >50 km. Para los viajes con origen o

destino fuera de la zonificación de estudio, se proporciona la distancia entre el origen/destino a nivel nacional y el punto de salida/entrada del país.

- f. 'actividad\_origen': tipo de actividad realizado en origen, diferenciando entre casa, trabajo\_estudio, otras actividades frecuentes ('frecuente'), y actividades esporádicas ('no\_frecuente')
  - g. 'actividad\_destino': tipo de actividad realizada en destino, diferenciando entre casa, trabajo\_estudio, otras actividades frecuentes ('frecuente'), y actividades esporádicas ('no\_frecuente')
  - h. 'estudio\_origen\_posible': toma el valor 'sí' si la actividad en origen puede hacer referencia a una actividad de tipo educativa. Toma el valor 'no' en caso contrario.
  - i. 'estudio\_destino\_posible': toma el valor 'sí' si la actividad en destino puede hacer referencia a una actividad de tipo educativa. Toma el valor 'no' en caso contrario
  - j. 'residencia': lugar de residencia a nivel provincial, según codificación INE.
  - k. 'renta': nivel de renta en miles de euros de los viajeros en los siguientes rangos: < 10, entre 10 y 15 y > 15
  - l. 'edad': rango etario de los viajeros ('0-24', '25-44', '45-64' y '>65'). Toma el valor 'NA' cuando dicha información no se facilita por criterios de privacidad<sup>5</sup>
  - m. 'sexo': sexo de los viajeros ('hombre' y 'mujer'). Toma el valor 'NA' cuando dicha información no se facilita por criterios de privacidad<sup>5</sup>
  - n. 'viajes': número de viajes expandidos
  - o. 'viajes\_km': producto de 'viajes' por la distancia ortodrómica de los viajes. Para los viajes con origen o destino fuera de la zonificación de estudio, se contabiliza solamente la distancia en territorio nacional.
3. **Viajes por persona.** Ficheros en formato 'CSV' con la información del número de viajes realizados por persona. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es 'AAAAMMDD\_Personas\_dia\_[zona].csv' donde 'AAAAMMDD' es el año, mes y día de estudio y '[zona]' el tipo de zonificación empleada pudiendo tomar los valores 'distritos', 'municipios' y 'GAU'. El fichero presenta los siguientes campos:
- a. 'fecha': fecha de estudio en formato AAAAMMDD
  - b. 'zona\_pernoctacion': identificador de la zona donde pernocta
  - c. 'edad': rango etario de los viajeros ('0-24', '25-44', '45-64' y '>65').
  - d. 'sexo': sexo de los viajeros ('hombre' y 'mujer').
  - e. 'número de viajes': número de viajes realizados por la personas en los grupos de '0' viajes, '1' viaje, '2' viajes y más de dos viajes ('2+' ).

---

<sup>5</sup> Solo se facilita información cuando el número de viajes es superior o igual a 10 unidades o si el marco muestral asociado a los estratos sociodemográficos es superior a 2.500 personas.

- f. 'personas': número de personas<sup>6</sup>
4. **Pernoctaciones.** Ficheros en formato 'CSV' con la información del número de personas que pernoctan en cada zona. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura del fichero es 'AAAAMMDD\_Pernoctaciones\_[zona].csv.gz' donde 'AAAAMMDD' es el año, mes y día de estudio y '[zona]' el tipo de zonificación empleada pudiendo tomar los valores 'distritos', 'municipios' y 'GAU'. El fichero presenta los siguientes campos:
- a. 'fecha': fecha de estudio en formato AAAAMMDD
  - b. 'zona\_residencia': lugar de residencia a nivel de distrito censal, según codificación INE
  - c. 'zona\_pernoctacion': identificador de la zona donde pernocta
  - d. 'personas': número de personas
5. **Zonas descartadas.** Ficheros en formato 'CSV' con la información de las zonas según la zonificación de 'distritos' descartadas por falta de muestra en el día de estudio. El fichero se denomina 'AAAAMMDD\_distritos\_descartados.txt' donde 'AAAAMMDD' es el año, mes y día de estudio. El fichero presenta los siguientes campos:
- a. 'fecha': fecha de estudio en formato AAAAMMDD
  - b. 'zona': identificador de la zona descartada

---

<sup>6</sup> Por motivos de privacidad, valores inferiores a 10 no son proporcionados

## Anexo I. Consideraciones metodológicas con respecto al análisis de movilidad durante el estado de alarma

Los estudios de movilidad diaria del presente proyecto pretenden dar continuidad a los estudios de monitorización continua de la movilidad llevados a cabo en el marco del proyecto 'Análisis de la movilidad en España con tecnología big data durante el Estado de Alarma para la gestión de la crisis del COVID-19'.

La metodología empleada para el cálculo de indicadores en este proyecto es muy similar a la utilizada en el anterior proyecto incorporando un conjunto de mejoras relevantes. Entre las mejoras incorporadas cabe destacar el cambio de la fuente de datos principal utilizada, siendo en este proyecto datos procedentes de sondas de red. Estos datos proporcionan una granularidad temporal significativamente superior a los datos de CDRs, permitiendo potencialmente una mejor caracterización de la movilidad (por ejemplo, detectando algunos viajes que con la fuente de datos anterior podían no reportarse por falta de registros) y un aumento sustancial de la muestra útil de usuarios, lo que disminuye el error muestral en los procesos de expansión de la muestra. Estas modificaciones metodológicas pueden motivar cambios en las estimaciones de los indicadores de movilidad entre ambos estudios. Por lo tanto, en el caso de que se quieran analizar ambas fuentes de información de manera conjunta, es importante tener en consideración que las variaciones de los valores de movilidad observados pueden venir motivados por: (i) el propio cambio de los patrones de movilidad desde 2021 y (ii) las modificaciones metodológicas incorporadas en este proyecto. Puede ser de utilidad analizar tendencias de variación de la movilidad dentro de cada serie de datos, con el objetivo de minimizar las diferencias motivadas por los diferentes planteamientos metodológicos empleados (por ejemplo, variación de la movilidad entre enero de 2021 y febrero de 2021 frente a la variación entre enero de 2022 y febrero de 2022).



