



Secretaría de Estado de Infraestructuras, Transporte y Vivienda

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

Informe metodológico

Versión 3 - 1 de febrero de 2024

Índice

1. INTRODUCCIÓN	3
1.1 ANTECEDENTES	3
1.2 OBJETO Y ALCANCE DEL ESTUDIO	4
1.3 GLOSARIO DE TÉRMINOS	7
2. FUENTES DE DATOS UTILIZADAS.....	9
2.1 REGISTROS ANONIMIZADOS DE TELEFONÍA MÓVIL	9
2.2 USOS DEL SUELO.....	9
2.3 DATOS DE POBLACIÓN	10
2.4 DATOS DE LA RED DE TRANSPORTE Y OFERTA Y DEMANDA DE SERVICIOS.....	10
2.5 DATOS DE DEMANDA DE TRANSPORTE.....	12
3. SOLUCIÓN TÉCNICA Y METODOLOGÍA.....	14
3.1 EXTRACCIÓN DE LOS REGISTROS DE TELEFONÍA MÓVIL.....	14
3.2 GENERACIÓN DE LOS INDICADORES DE MOVILIDAD	14
3.3 FIABILIDAD DE LOS RESULTADOS Y ERROR MUESTRAL	17
4. ENTREGABLES.....	18
4.1 ZONIFICACIÓN.....	18
4.2 ESTUDIOS BÁSICOS	18
4.3 ESTUDIOS COMPLETOS	20
4.4 ESTUDIOS DE RUTAS POR CARRETERA	27
ANEXO I. CONSIDERACIONES METODOLÓGICAS CON RESPECTO AL ANÁLISIS DE MOVILIDAD DURANTE EL ESTADO DE ALARMA	29

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
3	01/02/2024	Se incluye la información referente a los estudios completos. Se presenta la información de los entregables en tablas para facilitar su lectura.	1, 2, 3, 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 geocalizados 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”**¹, 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”**², 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.

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

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

1.2 Objeto y alcance del estudio

El objeto del proyecto es analizar la movilidad de viajeros a nivel nacional.

Se generarán tres tipos de estudios:

1. Estudio básico diario
2. Estudio completo (mensual)
3. Estudio de rutas por carretera

El alcance y contenido de cada tipo de estudio se detalla a continuación.

1.2.1 Estudio básico diario

Dando continuidad al estudio diario realizado en el 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”, se realiza un **estudio básico diario**, de acuerdo a las siguientes especificaciones:

- **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 se 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**³: se segmentan los viajes de acuerdo a las siguientes características de los viajeros:
 - Lugar de residencia a nivel provincial
 - Edad, según las siguientes franjas etarias: 0-24, 25-44, 45-64 y ≥ 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.

1.2.2 Estudio completo

Se realiza un **estudio completo con periodicidad mensual**, con un mayor nivel de detalle que el estudio básico diario, de acuerdo a las siguientes especificaciones:

- **Viajes objeto de estudio**: viajes de más de 500 metros con origen y/o destino en el territorio nacional excluyendo viajes realizados por conductores profesionales (conductores de vehículos de mercancías, repartidores, taxistas, tripulaciones de avión, etc.).
- **Población de estudio**: residentes en España y visitantes con residencia en el extranjero.
- **Período de estudio**: la duración del periodo de estudio es la misma que la duración de los estudios básicos. Para cada mes de estudio, se analiza una semana tipo y un conjunto de hasta 10 días singulares.
- **Zonificación**: se emplea la misma zonificación que en el estudio básico.
- **Identificación de etapas y modo de transporte**: los viajes cuya distancia ortodrómica entre origen y destino es superior a 5 km se segmentan en las etapas que los componen y se

³ Algunas de las variables sociodemográficas puede que no se proporcionen en algunos casos por criterios de privacidad; presentando el valor “NA” en tal caso.

identifica el modo de transporte de las cada etapa distinguiendo los siguientes modos: (i) carretera, diferenciando entre viajes en vehículo privado y autobús en el caso de viajes con origen y destino en Comunidades Autónomas diferentes; (ii) ferrocarril, distinguiendo entre alta velocidad y ferrocarril convencional; (iii) transporte aéreo; y (iv) transporte marítimo.

- **Resolución temporal:** los viajes y las etapas 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 viaje/etapa. 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 la distancia de cada viaje/etapa como la distancia entre el origen y el destino a lo largo de la ruta seguida sobre las infraestructuras de transporte y se emplean los siguientes intervalos: 500 metros-2 km, 2-5 km, 5-10 km, 10-50 km y >50 km. Para el caso de viajes al/desde el extranjero, la distancia se corresponde a la distancia entre el origen/destino del viaje y el punto de salida/entrada en el territorio nacional
- **Tiempo de viaje:** se determina el tiempo entre el inicio y el fin de cada etapa y del viaje en su conjunto.
- **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.
- **Segmentación por perfil sociodemográfico⁴:** los viajes/etapas de los residentes en España se segmentan de acuerdo a las siguientes características de los viajeros:
 - Lugar de residencia a nivel provincial.
 - Edad, según las siguientes franjas etarias: 0-24, 25-44, 45-64 y ≥ 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.
 - País de residencia.

Para cada estudio completo mensual se generarán los siguientes productos:

- **Matriz de etapas**, que proporciona los datos de viajes desagregados por etapas, indicando el modo de transporte de cada etapa. La matriz de etapas está ligada de manera unívoca a las matrices de viajes y de tours de manera que se pueda identificar las etapas y modos que componen cada conjunto de viajes o tours.
- **Matriz de viajes completa**, que además de la información contenida en la matriz de viajes básica, incluye la segmentación por modo de transporte, el número de conexiones o transbordos, y el tiempo de viaje.
- **Matriz de tours**, en la que se identifican las secuencias de viajes que componen tours, segmentados según los modos de transporte utilizados en dichas secuencias de viajes.
- **Matriz de movilidad transfronteriza**, que proporciona información sobre los viajes desde/hacia el extranjero indicando de la zona fronteriza nacional que se atraviesa.

⁴ Algunas de las variables sociodemográficas puede que no se proporcionen en algunos casos por criterios de privacidad; presentando el valor "NA" en tal caso.

- **Matrices del área de influencia de las principales terminales de transporte público**, que proporcionan información sobre los viajes que hacen uso de dichas terminales, caracterizando los modos de acceso y dispersión.
- **Matriz de movilidad obligada recurrente**, que proporciona información sobre la frecuencia de los viajes en función del par origen destino.

Estudios de rutas por carretera

Se lleva a cabo un estudio de las **rutas seguidas en los desplazamientos por carreteras y el total de viajeros que utilizan cada ruta**, de acuerdo a las siguientes especificaciones:

- **Viajes objeto de estudio**: viajes intermunicipales de más de 5.000 metros con origen y/o destino en el territorio nacional.
- **Población de estudio**: residentes en España y visitantes con residencia en el extranjero.
- **Período de estudio**: se analiza una semana tipo de invierno y una semana tipo de verano a definir conjuntamente con el MITMA.
- **Zonificación**: se proporcionan los datos de ruta segmentados de acuerdo a sus orígenes y destinos, según la zonificación empleada en el estudio básico y el estudio completo.
- **Segmentación por ruta utilizada y distancia de trayecto**: se distingue entre las diferentes rutas de carretera utilizadas, de acuerdo a la información georreferenciada de la red de carreteras.

Se generan los siguientes productos:

- **Geometría de tramos**, que proporciona la geometría de cada tramo de la red de carreteras.
- **Relación tramo-ruta**, que proporciona el conjunto de tramos que conforman cada ruta.
- **Información de tramos**, que proporciona el número de viajeros que emplean cada tramo de la red de carreteras.
- **Matrices de rutas**, que proporcionan el número de viajeros que emplean una ruta concreta.

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

Actividad	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.
Estancia	Permanencia de un individuo durante cierto tiempo en un lugar determinado.
Etapas	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 billeteaje, en el caso de que los dos trayectos se realicen mediante servicios distintos y se disponga del billeteaje de cada uno por separado.
Tour	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).
Viaje	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**⁵: 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.

⁵ 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 para los residentes en España y datos de las encuestas FRONTUR para los visitantes extranjeros.

2.4 Datos de la red de transporte y oferta y demanda de servicios

La solución técnica integra datos de la red de transporte e información sobre la oferta de servicios que resultan fundamentales para la determinación del modo de transporte y la ruta de los viajes. A continuación se describen las fuentes de datos empleadas para los diferentes modos de transporte.

2.4.1 Transporte por carretera

Las fuentes de datos empleadas para obtener de transporte por carretera se pueden dividir en dos grandes grupos: (i) la información asociada a la red de carreteras del territorio nacional y (ii) la información asociada a los servicios de transporte por autobús entre comunidades autónomas.

En relación a la información de la red de carreteras, ésta se obtiene mediante el uso de las APIs de OpenStreetMap (www.openstreetmap.org), obteniendo información de las distintas opciones de rutas para desplazarse en vehículo privado desde un punto 'A' hasta un punto 'B' del territorio. La información se obtiene considerando diferentes opciones para el cálculo de rutas (por ejemplo, sin restricciones, evitando peajes, evitando carreteras principales, etc.), de forma que se obtenga un árbol de rutas lo suficientemente amplio.

Por otro lado, para la información de viajes en autobús, se dispone de los datos de oferta de transporte asociado a las expediciones de contratos VACs (Viajeros, Autorización y Concesión), que contiene información trimestral sobre las rutas de autobús interurbanas asociadas a concesiones estatales. Estos datos no son públicos y han sido proporcionados por el MITMA para su uso en el marco de este proyecto. Los datos están disponibles en formato Excel y contienen, para cada uno de los VACs, la secuencia ordenada de paradas que recorren los autobuses en cada expedición. Para todas las paradas se dispone de un código identificador único, del nombre de la parada, de su dirección y del municipio en el que se ubica. Además, para una alta proporción de las paradas se dispone también de su longitud y de su latitud, lo que permite ubicarlas con precisión en el territorio.

2.4.2 Transporte ferroviario

Las fuentes de datos empleadas para obtener los datos de oferta de transporte ferroviario son las siguientes:

- **Información de ubicación de estaciones y oferta de servicios regulares de ferrocarriles RENFE AVLD** (media distancia, larga distancia y AVE), **RENFE Cercanías**, **RENFE FEVE**, **FGC**, **FGV**, **Euskotren** y **OUIGO**. Se dispone de información mensual de salidas y llegadas de servicios ferroviarios realizados en territorio nacional desagregada a nivel horario, así como de la ubicación de las estaciones. Aunque estos datos son públicos para algunos periodos y se pueden descargar del Punto de Acceso Nacional (NAP) de datos del transporte, en el marco del proyecto se dispone de acceso a datos históricos del NAP (facilitado por el MITMA), de

manera que los datos que se utilizan son siempre los más cercanos al mes de estudio. Los ficheros de servicios regulares de ferrocarriles RENFE AVLD, Cercanías, FEVE, FGC, FGV, Euskotren y OUIGO están en formato GTFS (General Transit Feed Information) y proporcionan información de los servicios mediante varios ficheros CSV. En particular, se dispone de un fichero con información georreferenciada de las estaciones de ferrocarril y de un fichero que describe la secuencia ordenada de paradas que se recorre en cada expedición ferroviaria.

- **Datos de la infraestructura de la red ferroviaria**, que contienen información georreferenciada de tramos y dependencias ferroviarias. Estos datos no son públicos y han sido proporcionados por ADIF exprofeso para la ejecución del proyecto. Los datos de la infraestructura ferroviaria se proporcionan en formato ESRI GeoDataBase y contienen la geometría de los tramos ferroviarios y de las dependencias, así como las relaciones entre las geometrías y los códigos identificadores.
- **Información de circulaciones de ferrocarriles**, en el que se especifica la secuencia ordenada de tramos que se recorre entre cada par de estaciones de origen y destino para servicios de cercanías y de larga y media distancia. Estos datos no son públicos y fueron proporcionados por ADIF exprofeso para la ejecución del proyecto. El fichero de circulaciones de ferrocarriles está en formato Excel y describe, para una fecha dada, la secuencia ordenada de tramos de ferrocarril que recorre un tren entre una estación de origen y una estación de destino. En el mismo fichero Excel se proporciona, además, la relación entre tramos y dependencias ferroviarias.

2.4.3 Transporte aéreo

La fuente de datos empleada para obtener información de oferta de transporte aéreo es la información facilitada por la AECFA (Asociación Española para la Coordinación y Facilitación de Franjas Horarias), que contiene información mensual de salida y llegada de vuelos realizados en territorio nacional y desagregada a nivel horario. Para algunos periodos, estos datos son públicos y se pueden descargar desde el NAP, pero en el marco del proyecto se dispone de acceso a datos históricos, de modo que los datos empleados en este proyecto son siempre los más cercanos al mes de estudio. Los datos están en formato GTFS, proporcionando un conjunto de ficheros CSV entre los que se encuentra un fichero que contiene información sobre las terminales y un fichero que describe la secuencia ordenada de paradas que se recorre en cada expedición.

2.4.4 Transporte marítimo

La fuente de datos principal empleada para obtener la oferta de transporte marítimo procede de la Demanda de Puertos del Estado, que contiene la información mensual de operaciones de entrada y salida a puertos nacionales, con el número de pasajeros embarcados y desembarcados. Estos datos no son públicos y han sido proporcionados por el MITMA para su uso en el marco de este proyecto.

También se cuenta con la información de los servicios regulares de ferry de las empresas Fred Olsen y Baleària, que operan en las Islas Canarias y en las Islas Baleares, respectivamente. Estos ficheros contienen información mensual de salida y llegada de servicios en ferry desagregados a nivel horario. Para algunos periodos, estos datos son públicos y se pueden descargar desde el NAP, pero en el marco del proyecto se dispone de acceso a datos históricos, de modo que los datos

empleados en este proyecto son siempre los más cercanos al mes de estudio. Los ficheros de servicios regulares de ferry de las empresas Fred Olsen y Baleària están en formato GTFS y proporcionan información de los servicios a través de ficheros CSV, de manera similar a los datos de transporte aéreo y ferroviario. En particular, se dispone de un fichero con información georreferenciada de puertos y de un fichero que describe la secuencia ordenada de puertos que se recorre en cada expedición de ferry.

2.5 Datos de demanda de transporte

Por último, la plataforma analítica de Nommon integra información procedente de otras fuentes de datos de demanda de transporte que se utiliza para ajustar los resultados de reparto modal estimados a partir de los datos de telefonía móvil para los viajes de distancia superior a 5 km.

Adicionalmente, los datos de demanda se han empleado en este proyecto para validar los datos de oferta de transporte y, en caso de ser necesario, completar la oferta para estar alineada con la demanda.

2.5.1 Transporte por carretera

La información de demanda diaria de autobús se obtiene fusionando los datos trimestrales de demanda procedentes de SITRANBUS, los datos mensuales de la Estadística de Transporte de Viajeros⁶ del Instituto Nacional de Estadística (INE) y los indicadores diarios de movilidad calculados en los estudios básicos agregados a nivel provincial.

Los datos procedentes de SITRANBUS contienen información trimestral acerca del número de viajeros que se desplazan de un municipio de origen a un municipio de destino en servicios de autobús asociados a concesiones estatales. Estos datos no son públicos y han sido proporcionados por el MITMA para su uso en el marco de este proyecto.

Los datos publicados por el INE en la Estadística de Transporte de Viajeros indican el volumen de viajeros mensuales en transporte interurbano por autobús regular divididos en cercanías, media distancia y larga distancia. A partir del número de viajeros de cada mes del trimestre por rangos de distancia, el volumen de viajeros trimestrales procedente de SITRANBUS se divide entre los tres meses que conforman el trimestre.

Finalmente, el volumen mensual estimado para cada par origen y destino de municipios se reparte a lo largo del mes en función del número de viajes diario entre la provincia de origen y destino calculado en los estudios básicos.

2.5.2 Transporte ferroviario

Para el ajuste del modo ferroviario, se cuenta con la demanda de los trayectos considerados Obligación de Servicio Público (OSP) de RENFE que contiene información diaria sobre el número de pasajeros entre estaciones. Los servicios públicos de RENFE engloban Media Distancia, Ancho métrico y AVANT. Estos datos no son públicos y han sido proporcionados por el MITMA para su uso en el marco de este proyecto. Debido a que existen pares de estaciones conectadas tanto por

⁶ <https://www.ine.es/jaxiT3/Tabla.htm?t=20239>

servicios públicos como por servicios comerciales, para los que no se ha dispuesto del valor de demanda, en el presente proyecto sólo se ha empleado el dato de demanda de aquellos pares de estaciones que únicamente están conectadas por servicios públicos.

2.5.3 Transporte aéreo

La fuente de datos empleada para obtener información de demanda de transporte aéreo es la información facilitada por AENA, que contiene información diaria sobre las operaciones realizadas en aeropuertos nacionales indicando para cada una de estas operaciones si se trata de una operación de llegada/salida, el aeropuerto de origen/destino del viaje, la hora en que se realiza la operación y el número de pasajeros. Estos datos no son públicos y han sido proporcionados por el MITMA para su uso en el marco de este proyecto.

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”

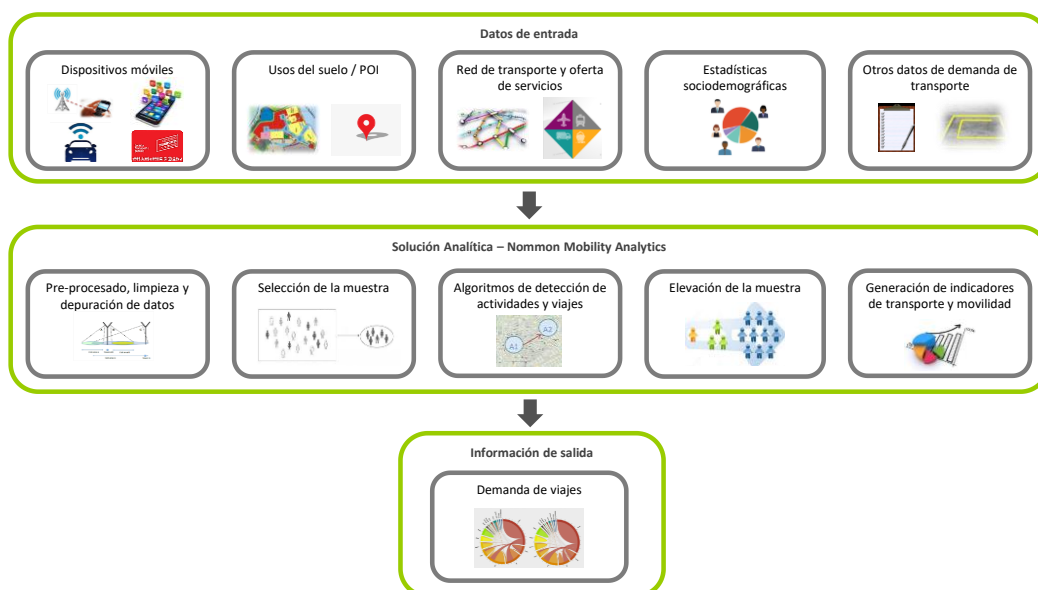


Figura 1 – Esquema de alto nivel de la solución técnica empleada en el proyecto

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). Para la identificación de modo de transporte y ruta se emplean técnicas de *map matching* que tienen en consideración criterios basados en el tiempo de desplazamiento, velocidad, topología de la red de transporte y compatibilidad con la oferta de servicios de transporte, permitiendo obtener una métrica de similitud para cada una de las alternativas de transporte. A cada desplazamiento se le asigna *a priori* la alternativa de transporte que maximiza la probabilidad de aparición de los registros observados (p.ej., un conjunto de registros a lo largo de las vías ferroviarias de alta velocidad, en un horario y tiempos de viaje compatible con los

servicios existentes, corresponderá a una etapa en ferrocarril de alta velocidad). 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), hora de finalización (hora de inicio de la actividad siguiente) y la secuencia de etapas que lo componen detallando el modo y ruta empleada.

5. **Elevación de la muestra.** Una vez construidos los diarios de actividades y viajes con todos sus atributos, es necesario elevar la muestra de usuarios al total de la población objeto de estudio. Cabe distinguir dos casos diferenciados: la elevación de la muestra de residentes en España y la elevación de la muestra de no residentes. En el caso de los residentes en España, 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. En el caso de los no residentes, la expansión de la muestra se realiza a partir de los datos de la Estadística de Movimientos Turísticos en Fronteras (FRONTUR) elaborada por el INE aplicando factores de expansión estratificados por país de residencia y duración de la estancia.
6. **Ajuste con información de demanda de referencia cierta.** Los resultados obtenidos a partir de los datos de telefonía móvil y su fusión con otras fuentes de datos son ajustados utilizando datos de demanda de referencia cierta de avión, tren y autobús. La solución empleada se basa en un ajuste a nivel desagregado, evaluando el grado de bondad en la estimación de los modos de transporte de cada viaje a partir de las técnicas descritas en el punto 4 y ajustando los resultados teniendo en consideración dicho grado de bondad.
7. **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 la sección 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:

4.1 Zonificación

En el proyecto, aparte de la zonificación principal de estudio (3.909), 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:

- **“zonificacion_distritos.shp”**, zonificación de mayor detalle compuesta por distritos o agregaciones de distritos.
- **“zonificacion_municipios.shp”**, agregación de la zonificación de detalle a nivel municipal o agregaciones de municipios.
- **“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.

4.2 Estudios básicos

4.2.1 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:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
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.
origen	Identificador de la zona de origen. Para los viajes con origen fuera de la zonificación de estudio, se define el origen "externo"
destino	Identificador de la zona de destino. Para los viajes con destino fuera de la zonificación de estudio, se define el destino "externo"
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 el valor de la distancia recorrida en territorio nacional
actividad_origen	Tipo de actividad realizada en la zona de origen del viaje, diferenciando entre: - "casa" - "trabajo_estudio" - "frecuente" - "no_frecuente"
actividad_destino	Tipo de actividad realizada en la zona de destino del viaje, diferenciando entre: - "casa" - "trabajo_estudio" - "frecuente" - "no_frecuente"
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
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
residencia	Lugar de residencia a nivel provincial, según codificación INE.
renta	Nivel de renta (en miles de euros) de los viajeros en los siguientes rangos: - <10: menos de 10,000 euros. - 10-15: entre 10,000 y 15,000 euros. - >15: más de 15,000 euros.
edad	Rango etario de los viajeros: - 0-24: 0-24 años - 25-44: 25-44 años - 45-64: 45-64 años - >65: > 65 años Toma el valor "NA" cuando la información no se pueda proporcionar por motivos de privacidad.
sexo	Toma el valor "NA" cuando la información no se pueda proporcionar por motivos de privacidad.
viajes	Número de viajes expandidos
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

4.2.2 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:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
zona_pernoctacion	Identificador de la zona de pernoctación
edad	Rango etario de los viajeros: - 0-24: 0-24 años - 25-44: 25-44 años - 45-64: 45-64 años - >65: > 65 años Toma el valor "NA" cuando la información no se pueda proporcionar por motivos de privacidad.
sexo	Sexo de los viajeros: 'hombre' y 'mujer' Toma el valor "NA" cuando la información no se pueda proporcionar por motivos de privacidad.
número de viajes	Número de viajes realizados por la personas en los grupos de '0', '1', '2' y '2+'
personas	Número de personas.

4.2.3 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:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD.
zona_residencia	Identificador de la zona de residencia.
zona_pernoctacion	Identificador de la zona de pernoctación.
personas	Número de personas.

4.2.4 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:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD.
zona	Identificador de la zona.

4.3 Estudios completos

4.3.1 Matrices de etapas

Ficheros en formato "CSV" con la información de las etapas de los viajes segmentadas por modo y medio de transporte. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es "AAAAMMDD_Etapas_[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 etapas presenta los siguientes campos:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD.
tipo_dia	Código que identifica el día de la semana y si es un día que forma parte de una semana estándar de estudios completos (día "tipo") o si es un día que se estudia porque su movilidad presenta patrones especiales (día "especial"), como es el caso de días festivos, vísperas de puentes, etc. Puede tomar los valores "LT", "MT", "XT", "JT", "VT", "ST", "DT" (Lunes Tipo, Martes Tipo, Miércoles Tipo, etc.) para días estándar o "LE", "ME", "XE", "JE", "VE", "SE", "DE" (Lunes Especial, Martes Especial, Miércoles Especial, etc.) para días especiales.
ID	Identificador único de etapa para cada combinación de zona de origen, zona de destino, periodo horario, modo, y medio. Este código se usa para relacionar las matrices de viajes completas y la de etapas.
periodo	Hora de origen de la etapa, en formato "HH".
origen	Identificador de la zona de origen.
destino	Identificador de la zona de destino.
modo	Modo de transporte empleado, diferenciando entre "C" (modo de carretera), "F" (modo ferroviario), "A" (modo aéreo) y "M" (modo marítimo). El campo "modo" puede tomar el valor "NA" para los viajes de menos de 5 km que no tengan origen y destino dentro de una zona urbana a estudiar con detalle mediante encuestas.
medio	Medio de transporte empleado: - En el caso de que el valor del campo "modo" sea "A" o "M", el valor de este campo será "A" o "M", respectivamente. - En el caso de que el valor del campo "modo" sea "F", el valor de este campo podrá ser "FC" (ferrocarril convencional) o "AV" (ferrocarril de alta velocidad). El tipo de servicio "convencional" incluye los servicios Intercity, Media Distancia, Regional, Regional Express y Euromed; el tipo de servicio "alta velocidad" incluye los servicios ALVIA, AVANT, AVCity, AVE, AVE-INT, AVE-TGV, AVLO y los servicios de las compañías Ouigo e Iryo. - En el caso de que el valor del campo "modo" sea "C", el valor de este campo podrá ser "PR" (vehículo privado) o "AU" (autobús). El medio de autobús solo se distingue en viajes entre comunidades autónomas. El valor del campo "medio" es "NA" para las etapas en las que el campo "modo" también sea igual a "NA".
distancia	Distancia entre el origen y el destino medida en km a lo largo de la ruta seguida sobre las infraestructuras de transporte. En caso de que el modo sea "NA", la distancia se corresponde con la distancia en línea recta entre el origen y el destino. El campo "distancia" puede tomar los siguientes valores: - 0.5-2: viajes de entre 0.5 km y 2 km. - 2-5: viajes de entre 2 km y 5 km. - 5-10: viajes de entre 5 km y 10 km. - 10-50: viajes de entre 10 km y 50 km. - >50: viajes de más de 50 km.
duracion	Tiempo de recorrido de la etapa. El campo "duracion" puede tomar los siguientes valores: - 0-1: viajes de menos de 1 hora. - 1-2: viajes de entre 1 hora y 2 horas. - 2-3: viajes de entre 2 horas y 3 horas. - 3-4: viajes de entre 3 horas y 4 horas. - >4: viajes de más de 4 horas.
pais	Código identificador del país o grupo de países de residencia del viajero. El campo "pais" puede tomar los siguientes valores: - ES (España) - DE (Alemania)

	<ul style="list-style-type: none"> - BE (Bélgica) - FR (Francia) - IE (Irlanda) - IT (Italia) - NL (Países Bajos) - PNO (Países Nórdicos) - PT (Portugal) - GB (Reino Unido) - REU (Resto de Europa) - RMU (Resto del Mundo)
residencia	Lugar de residencia a nivel provincial, según codificación INE.
renta	<p>Nivel de renta (en miles de euros) de los viajeros en los siguientes rangos:</p> <ul style="list-style-type: none"> - <10: menos de 10,000 euros - 10-15: entre 10,000 y 15,000 euros - >15: más de 15,000 euros <p>Toma el valor 'NA' para los viajeros residentes extranjeros.</p>
edad	<p>Rango etario de los viajeros:</p> <ul style="list-style-type: none"> - 0-24: 0-24 años - 25-44: 25-44 años - 45-64: 45-64 años - >65: > 65 años <p>Toma el valor 'NA' para los viajeros residentes extranjeros o cuando la información no se pueda proporcionar por motivos de privacidad.</p>
sexo	<p>Sexo de los viajeros; puede tomar los valores "hombre" o "mujer".</p> <p>Puede tomar el valor 'NA' para los viajeros residentes extranjeros o cuando la información no se pueda proporcionar por motivos de privacidad.</p>
viajes	Número de etapas expandidas, sin considerar a los transportistas profesionales.
viajes_km	Producto del campo "viajes" por la distancia de la etapa.

4.3.2 Matrices de viajes completas

Ficheros en formato "CSV" con la información de movilidad segmentada por modo y medio de transporte. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es "AAAAMMDD_Viajes_completa_[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:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
tipo_dia	<p>Código que identifica el día de la semana y si es un día que forma parte de una semana estándar de estudios completos (día "tipo") o si es un día que se estudia porque su movilidad presenta patrones especiales (día "especial"), como es el caso de días festivos, vísperas de puentes, etc.</p> <p>Puede tomar los valores "LT", "MT", "XT", "JT", "VT", "ST", "DT" (Lunes Tipo, Martes Tipo, Miércoles Tipo, etc.) para días estándar o "LE", "ME", "XE", "JE", "VE", "SE", "DE" (Lunes Especial, Martes Especial, Miércoles Especial, etc.) para días especiales.</p>
ID	<p>Identificador único de viaje para cada combinación de zona de origen, zona de destino, periodo horario, modo ppal y medio ppal.</p> <p>Este código se usa para relacionar la matriz de viajes con las matrices de tours, de influencia de terminales de transporte y de movilidad transfronteriza.</p>
periodo	Hora de origen del viaje, en formato "HH"
itinerario	<p>Secuencia ordenada de las etapas que componen el viaje, concatenadas por puntos y comas ";".</p> <p>Se hace referencia a cada etapa mediante el valor del campo "ID" de la matriz de etapas asociada "AAAAMMDD_Viajes_etapas_distritos.csv.gz".</p>
conexiones	Número de conexiones/transbordos que componen el viaje.
origen	Identificador de la zona de origen.

destino	Identificador de la zona de destino.
actividad_origen	Tipo de actividad realizada en la zona de origen del viaje, diferenciando entre: - "casa" - "trabajo_estudio" - "frecuente" - "no_frecuente"
actividad_destino	Tipo de actividad realizada en la zona de destino del viaje, diferenciando entre: - "casa" - "trabajo_estudio" - "frecuente" - "no_frecuente"
modo_ppal	Modo de transporte empleado en la etapa de mayor distancia que compone el viaje, diferenciando entre "CCP" (modo de carretera realizado por un conductor profesional), "CPA" (modo de carretera realizado por un pasajero), "F" (modo ferroviario), "A" (modo aéreo) y "M" (modo marítimo). El campo "modo_ppal" puede tomar el valor "NA" para los viajes de menos de 5 km que no tengan origen y destino dentro de una zona urbana a estudiar con detalle mediante encuestas.
medio_ppal	Medio de transporte empleado en la etapa de mayor distancia que compone el viaje: - En el caso de que el valor del campo "modo_ppal" sea "A" o "M", el valor de este campo será "A" o "M", respectivamente. - En el caso de que el valor del campo "modo_ppal" sea "F", el valor de este campo podrá ser "FC" (ferrocarril convencional) o "AV" (ferrocarril de alta velocidad). - En el caso de que el valor del campo "modo_ppal" sea "C", el valor de este campo podrá ser "PR" (vehículo privado) o "AU" (autobús). El medio "autobús" solo se distingue en viajes entre comunidades autónomas. - En el caso de que el valor del campo "modo_ppal" sea "NA", el valor de este campo será "NA"
modos	Secuencia ordenada de modos de las etapas que componen el viaje. Es igual a la concatenación de los modos de cada una de las etapas que conforman el viaje, separados por guiones "-".
medios	Secuencia ordenada de modos de las etapas que componen el viaje. Es igual a la concatenación de los medios de cada una de las etapas que conforman el viaje, separados por guiones "-".
distancia	Distancia entre el origen y el destino medida en km a lo largo de la ruta seguida sobre las infraestructuras de transporte. En caso de que el campo "modo_ppal" sea "NA", la distancia se corresponde con la distancia en línea recta entre el origen y el destino. El campo "distancia" puede tomar los siguientes valores: - 0.5-2: viajes de entre 0.5 km y 2 km. - 2-5: viajes de entre 2 km y 5 km. - 5-10: viajes de entre 5 km y 10 km. - 10-50: viajes de entre 10 km y 50 km. - >50: viajes de más de 50 km.
duracion	Tiempo de viaje. El campo "duracion" puede tomar los siguientes valores: - 0-1: viajes de menos de 1 hora. - 1-2: viajes de entre 1 hora y 2 horas. - 2-3: viajes de entre 2 horas y 3 horas. - 3-4: viajes de entre 3 horas y 4 horas. - >4: viajes de más de 4 horas.
viajes	Número de viajes expandidos, sin considerar a los transportistas profesionales.
viajes_km	Producto del campo "viajes" por la distancia del viaje.

4.3.3 Matrices de tours

Ficheros en formato "CSV" con la información de tours basados en el lugar de pernoctación realizados durante el día de estudio. Se generan distintos ficheros en función de las distintas

zonificaciones. La nomenclatura de los ficheros es "AAAAMMDD_Tours_[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 tours presenta los siguientes campos:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
tipo_dia	Código que identifica el día de la semana y si es un día que forma parte de una semana estándar de estudios completos (día "tipo") o si es un día que se estudia porque su movilidad presenta patrones especiales (día "especial"), como es el caso de días festivos, vísperas de puentes, etc. Puede tomar los valores "LT", "MT", "XT", "JT", "VT", "ST", "DT" (Lunes Tipo, Martes Tipo, Miércoles Tipo, etc.) para días estándar o "LE", "ME", "XE", "JE", "VE", "SE", "DE" (Lunes Especial, Martes Especial, Miércoles Especial, etc.) para días especiales.
itinerario	Secuencia ordenada de los identificadores de viajes que componen el tour, concatenados por coma ",". Los identificadores de los viajes que aparecen en este campo hacen referencia al valor del campo "ID" en la matriz de viajes asociada "AAAAMMDD_Viajes_completa_distritos.csv.gz".
origen	Identificador de la zona de origen del tour. La zona de origen del tour será siempre el lugar de pernoctación de un viajero.
zonas	Secuencia ordenada de las zonas visitadas en el tour, concatenadas por comas ",".
modos_ppales	Secuencia ordenada de los modos principales de los viajes que componen el tour, concatenados por coma ",".
medios_ppales	Secuencia ordenada de los medios principales de los viajes que componen el tour, concatenados por coma ",".
distancias	Conjunto de distancias de los viajes que componen el tour. Es igual a la secuencia ordenada de las distancias de cada uno de los viajes del tour, concatenadas por coma ",".
duraciones	Conjunto de duraciones de los viajes que componen el tour. Es igual a la secuencia ordenada de las duraciones de cada uno de los viajes del tour, concatenadas por coma ",".
personas	Número de personas expandidas que realizan el tour, sin considerar a los transportistas profesionales.
personas_km	Producto del campo "personas" por la distancia de cada tour.

4.3.4 Matrices de movilidad transfronteriza

Ficheros en formato "CSV" con la información sobre los viajes desde/hacia el extranjero con indicación de la zona fronteriza nacional que se atraviesa. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es "AAAAMMDD_Transfronteriza_[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 movilidad transfronteriza presenta los siguientes campos:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
tipo_dia	Código que identifica el día de la semana y si es un día que forma parte de una semana estándar de estudios completos (día "tipo") o si es un día que se estudia porque su movilidad presenta patrones especiales (día "especial"), como es el caso de días festivos, vísperas de puentes, etc. Puede tomar los valores "LT", "MT", "XT", "JT", "VT", "ST", "DT" (Lunes Tipo, Martes Tipo, Miércoles Tipo, etc.) para días estándar o "LE", "ME", "XE", "JE", "VE", "SE", "DE" (Lunes Especial, Martes Especial, Miércoles Especial, etc.) para días especiales.
id_viaje	Identificador único del viaje que tiene origen y/o destino en una zona transfronteriza. Hace referencia al campo "ID" de la matriz de viajes completa "AAAAMMDD_Viajes_completas_distritos.csv.gz"
origen	Identificador de la zona de origen.

destino	Identificador de la zona de destino.
punto_entrada	Zona fronteriza de entrada a España. En el caso de viajes por carretera se indicará la zona fronteriza atravesada y, en el caso de otros modos, la estación/puerto/aeropuerto en la que termina la etapa con origen en el extranjero y destino en España. En caso de no tratarse de un viaje con origen en el extranjero, este campo toma el valor 'NA'.
punto_salida	Zona fronteriza de salida de España. En el caso de viajes por carretera se indicará la zona fronteriza atravesada y, en el caso de otros modos, la estación/puerto/aeropuerto en la que comienza la etapa al extranjero. En caso de no tratarse de un viaje con destino en el extranjero, este campo toma el valor 'NA'.
distancia	Distancia entre el origen y el destino medida en km a lo largo de la ruta seguida sobre las infraestructuras de transporte. En caso de que el campo "modo_ppal" sea "NA", la distancia se corresponde con la distancia en línea recta entre el origen y el destino. El campo "distancia" puede tomar los siguientes valores: - 0.5-2: viajes de entre 0.5 km y 2 km. - 2-5: viajes de entre 2 km y 5 km. - 5-10: viajes de entre 5 km y 10 km. - 10-50: viajes de entre 10 km y 50 km. - >50: viajes de más de 50 km.
viajes	Número de viajes expandidos, sin considerar a los transportistas profesionales.
viajes_km	Producto de 'viajes' por la distancia de los viajes.

4.3.5 Matrices del área de influencia de las principales terminales de transporte público

Ficheros en formato "CSV" con la información sobre los viajes que hacen uso de dichas terminales, caracterizando los modos de acceso y dispersión. Se generan distintos ficheros en función de las distintas zonificaciones. La nomenclatura de los ficheros es "AAAAMMDD_Influencia_TP_[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 movilidad transfronteriza presenta los siguientes campos:

Campos	Descripción
fecha	Fecha de estudio en formato AAAAMMDD
tipo_dia	Código que identifica el día de la semana y si es un día que forma parte de una semana estándar de estudios completos (día "tipo") o si es un día que se estudia porque su movilidad presenta patrones especiales (día "especial"), como es el caso de días festivos, vísperas de puentes, etc. Puede tomar los valores "LT", "MT", "XT", "JT", "VT", "ST", "DT" (Lunes Tipo, Martes Tipo, Miércoles Tipo, etc.) para días estándar o "LE", "ME", "XE", "JE", "VE", "SE", "DE" (Lunes Especial, Martes Especial, Miércoles Especial, etc.) para días especiales.
id_viaje	Identificador único de viaje que tiene alguna etapa con origen y/o destino en la terminal, siendo ese viaje un viaje del entregable "AAAAMMDD_Viajes_completas_GAU.csv.gz".
id_etapa	Identificador único de etapa del viaje que tiene origen y/o destino en la terminal, siendo esa etapa una etapa del entregable "AAAAMMDD_Viajes_etapas_distritos.csv.gz"
tipo_terminal	Tipo de terminal: - aeropuerto - puerto - tren - bus
estacion	Identificador del aeropuerto, puerto, estación de autobús o estación de ferrocarril. Se emplea la columna "Código estación" como identificador de las estaciones

terminal	Identificador de la terminal en aquellos casos en los que se divide la estación en terminales.
tipo_etapa	Relación que existe entre la etapa y la terminal de transporte considerada. Este campo puede tomar los siguientes valores: - "C" (conexión): si la etapa anterior que realiza el viajero es una etapa con destino en la terminal en el modo asociado a la terminal y la siguiente etapa que realiza el viajero es una etapa con origen en la terminal en el modo asociado a la terminal. - "A" (acceso): si la etapa no es una conexión y la siguiente etapa que realiza el viajero es una etapa con origen en la terminal en el modo asociado a la terminal. - "D" (dispersión): si la etapa no es una conexión y la etapa anterior que realiza el viajero es una etapa con destino en la terminal en el modo asociado a la terminal.
origen_viaje	Identificador de la zona de origen del viaje.
destino_viaje	Identificador de la zona de destino del viaje.
pais	Código identificador del país o grupo de países de residencia del viajero. El campo "pais" puede tomar los siguientes valores: - ES (España) - DE (Alemania) - BE (Bélgica) - FR (Francia) - IE (Irlanda) - IT (Italia) - NL (Países Bajos) - PNO (Países Nórdicos) - PT (Portugal) - GB (Reino Unido) - REU (Resto de Europa) - RMU (Resto del Mundo)
viajes	Número de viajeros expandidos, sin considerar a los transportistas profesionales.
viajes_km	Producto del valor del campo "viajes" por la distancia del viaje.

4.3.6 Matriz de movilidad obligada recurrente

Ficheros en formato "CSV" con la información sobre la frecuencia de los viajes para cada par OD en base a la zonificación de distritos. La nomenclatura de los ficheros es "AAAAMMDD_Movilidad_obligada_distritos.csv.gz", donde "AAAAMMDD" es el año, mes y día de estudio. La matriz de movilidad obligada presenta los siguientes campos:

Campos	Descripción
mes	Mes de estudio en formato AAAAMM
origen	Identificador de la zona de origen en base a "zonificacion_distritos".
destino	Identificador de la zona de destino en base a "zonificacion_distritos".
edad	Rango etario de los viajeros: - 0-24: 0-24 años - 25-44: 25-44 años - 45-64: 45-64 años - >65: > 65 años Puede tomar el valor 'NA' para los viajeros residentes extranjeros o cuando la información no se pueda proporcionar por motivos de privacidad.
sexo	Sexo de los viajeros; puede tomar los valores "hombre" o "mujer". Puede tomar el valor 'NA' para los viajeros residentes extranjeros o cuando la información no se pueda proporcionar por motivos de privacidad.
residencia	Lugar de residencia a nivel provincial, según codificación INE.
recurrencia	Número de días que los usuarios realizan un viaje de "origen" a "destino" durante las dos semanas del mes analizadas
personas	Número de personas expandidas que satisfacen la segmentación dada por los campos previos, sin considerar a los transportistas profesionales

4.4 Estudios de rutas por carretera

4.4.1 Matrices de rutas

Ficheros en formato "CSV" con la información de los viajes segmentados por ruta y distancia de viaje. La nomenclatura de los ficheros es "AAAAMMDD_OD_rutas.csv.gz", donde "AAAAMMDD" es el año, mes y día de estudio. La matriz de rutas presenta los siguientes campos:

Campos	Descripción
origen	Identificador de la zona de origen.
destino	Identificador de la zona de destino.
ruta	Identificador de ruta utilizada para desplazarse de origen a destino.
distancia	Distancia entre el origen y el destino medida en km a lo largo de la ruta seguida por carretera. El campo "distancia" puede tomar los siguientes valores: - 0.5-2: viajes de entre 0.5 km y 2 km. - 2-5: viajes de entre 2 km y 5 km. - 5-10: viajes de entre 5 km y 10 km. - 10-50: viajes de entre 10 km y 50 km. - >50: viajes de más de 50 km.
mediana_t	Duración mediana de los viajes bajo la segmentación indicada.
viajes	Número de viajeros expandidos, sin considerar a los transportistas profesionales.

4.4.2 Relación tramo-ruta

Ficheros en formato "CSV" con la información de la composición de rutas, indicando los tramos forman cada ruta. La nomenclatura de los ficheros es "AAAAMMDD_relacion_tramo_ruta.csv.gz", donde "AAAAMMDD" es el año, mes y día de estudio. El fichero relacional presenta los siguientes campos:

Campos	Descripción
ruta	Identificador de ruta.
tramo	Identificador de tramo.

4.4.3 Geometría de tramos

Ficheros en formato "SHP" con la información georreferenciada de los tramos utilizados en el estudio de rutas. La nomenclatura de los ficheros es "geometria_tramos.shp", donde se recogen todos los tramos utilizados en el estudio. El fichero presenta los siguientes campos:

Campos	Descripción
tramo	Identificador de tramo.
geometry	Geometría del tramo.

4.4.4 Información de tramos

Ficheros en formato "CSV" con información acerca de la naturaleza de los viajes que utilizan cada tramo. La nomenclatura de los ficheros es "AAAAMMDD_tramos_info.csv.gz", donde "AAAAMMDD" es el año, mes y día de estudio. La matriz de tramos presenta los siguientes campos:

Campos	Descripción
tramo	Identificador de tramo.

total	Número total de viajes que pasan por el tramo.
corto	Número total de viajes de menos de 50 km que pasan por el tramo.
medio	Número total de viajes de más de 50 km y menos de 250km que pasan por el tramo.
largo	Número total de viajes de más de 250 km que pasan por el tramo.
intra_provincial	Número total de viajes realizados con origen y destino en la misma provincia que pasan por el tramo.
inter_provincial	Número total de viajes realizados con origen y destino en diferente provincia que pasan por el tramo.
intra_ccaa	Número total de viajes realizados con origen y destino en la misma comunidad autónoma que pasan por el tramo.
inter_ccaa	Número total de viajes realizados con origen y destino en diferente comunidad autónoma que pasan por el tramo.

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).

