

# TAPTAP COVID-19 REPORT

## Geospatial Analysis

---

COUNTRY: SPAIN  
DATE: FEBRUARY 2020 - APRIL 2020

The logo for TAPTAP, featuring a white square icon with a stylized 'T' shape inside, followed by the word 'TAPTAP' in a bold, white, sans-serif font.

**TAPTAP**

## SOBRE TAPTAP

TAPTAP Digital es una multinacional española del sector marketing tecnológico fundada en 2010 con presencia en 10 países y ámbito global. TAPTAP, a través de su plataforma propietaria Sonata, ofrece soluciones en base a la inteligencia geoespacial aplicada a la segmentación y perfilación avanzada de audiencias, la activación publicitaria dinámica, y la medición de resultados offline y online.

Sonata Location Intelligence (LI) es el sistema de Sonata que permite el **análisis multivariable de fuentes de datos estáticas y dinámicas en tiempo real** para obtener índices e indicadores comportamentales, que permiten identificar y analizar patrones de los ciudadanos a nivel geográfico.

El ecosistema de datos de Sonata se basa en fuentes de datos anónimas estáticas y dinámicas, propias y de terceros, como la navegación móvil, datos de geoposición, datos censales estadísticos y otras fuentes dinámicas de datos en tiempo real. TAPTAP y sus colaboradores proporcionan un análisis independiente a partir de datos no identificativos que han sido suministrados y verificados dentro del marco legal comunitario europeo vigente basado en el consentimiento.

## ANTECEDENTES

En el contexto del confinamiento global, TAPTAP ha desarrollado un primer estudio para entender cuál es la relación entre la evolución de la pandemia y los hábitos de movilidad.

Para este estudio TAPTAP ha colaborado con la Universidad Carlos III de Madrid, incorporando expertos en la rama de Big Data e Inteligencia Artificial (IA), y las empresas Predicio y Tamoco, aportando datos de localización de alta frecuencia y precisión.

Este primer informe se centra en España desde que las restricciones en la movilidad fueron anunciadas, debido al virus conocido como COVID19, hasta mediados de Abril 2020. Al mismo tiempo, se desarrollan estudios multivariable por región con respecto a la movilidad permitida a partir de puntos de interés (PDI) de vulnerabilidad alta (hospitales, farmacias, residencias) de cara a poder determinar cuales son las regiones con mayor vulnerabilidad. Será necesario de un análisis continuado en el tiempo donde la relación causa-efecto no es siempre inmediata o directa y por ello publicaremos un informe adicional en las próximas semanas aportando mayor continuidad al análisis hasta la fecha.

Por último, creemos muy pertinente entender cuál es la vinculación de la vuelta a la actividad industrial/comercial y de consumo con respecto a la evolución de la pandemia en las próximas semanas. Por ello, la llegada de las fases de desconfinamiento será motivo de monitorización ya que podría acarrear nuevos brotes pandémicos poniendo en peligro la recuperación económica.

## SOBRE LOS DATOS

### Garantía de Privacidad:

Los datos de localización/movilidad han sido proporcionados por la tecnología Sonata propiedad de TAPTAP Digital SL. Sonata es una plataforma de marketing avanzado que proporciona acceso a datos de movilidad anonimizados, y recolectados con el debido consentimiento a través del marco legal proporcionado por el Reglamento General de Protección de Datos (RGPD). Los datos se recopilan de usuarios que han optado por proporcionar acceso a sus datos de ubicación y se han sometido a procesos que permiten la anonimización y evitan que los datos puedan vincularse a un individuo. Adicionalmente se ha contado con los datos de ubicación de gran precisión proporcionados por empresas como Predicio o Tamoco, quienes han optado por sumarse a este estudio. En todos los casos, se han tomado las medidas necesarias para garantizar que los datos tratados para la realización de este informe sean datos anonimizados. [Más información sobre los datos y la metodología.](#)

Para el estudio de movilidad, las muestras de datos procesadas provienen de los meses de Marzo y parte de Abril del 2020. El resultado final de esta muestra de alta representatividad y frecuencia contempla aproximadamente el 2% de la población total de España. Se han utilizado más de 300 millones de datos de localización durante un periodo de unas 6 semanas que cumplen con suficientes características de precisión y disponibilidad garantizando la relevancia y la representación.

## ANÁLISIS GENERAL

**Los índices multivariables** presentados, aportan una imagen compuesta de cada zona, público o tendencia. Permitiendo analizar el comportamiento del desarrollo de la pandemia Covid-19 en función del territorio, la evolución de la movilidad ciudadana con respecto a las medidas de confinamiento, los factores externos (e indicadores estáticos) que influyen en el desarrollo de la pandemia, la propagación pandémica en relación a la reactivación escalonada de la actividad industrial/comercial.

Se han identificado 4 principales bloques o conjuntos de variables para ordenar y englobar el estudio.

1. Datos oficiales de COVID19 (Casos, Hospitalizados, Recuperados, UCI, Fallecidos)
2. Comportamiento del usuario (Movilidad) Indicadores de **cambios de hábitos en la población** en el contexto COVID19.
3. Demografía y variables sociales: identificación del **riesgo de cada territorio** y los índices de vulnerabilidad naturales en función de la composición de sus habitantes.

4. Contexto Geográfico: Puntos y zonas de interés con alta demanda o saturación que pueden exponer a la población o condicionar sus hábitos. **Densidad de usuarios por área y por punto de interés.**

### CONSUMO POST COVID19

Estamos en un momento inédito, donde no existe una hoja de ruta, presentando una disyuntiva muy compleja para las instituciones. La garantía búsqueda de la salud personal de los ciudadanos a través del confinamiento significa la destrucción masiva de la economía, del empleo y del motor económico de cualquier país, la producción y el consumo. A estos efectos, desde TAPTAP, como aliado de grandes marcas e Instituciones para la ejecución de campañas de comunicación publicitaria basadas en datos, entendemos que los indicadores en comportamiento, tanto a nivel físico como digital de los consumidores, así como los propios del consumo, son claves para poder entender la relación de la reactivación paulatina de la economía con respecto a la evolución futura pandémica. De la misma manera, entendemos que existirá un nuevo consumidor post-COVID19, con caminos de compra alterados. Por lo tanto, a medida que produzcamos análisis complejos multivariable, que permiten correlacionar diferentes fuentes de datos, enriquecemos la visión del consumidor, sus hábitos, su relación con las marcas y las instituciones.

### CONCLUSIONES DEL ESTUDIO

**1.** Las medidas de confinamiento han logrado reducir la movilidad global en todas las provincias de España, aproximadamente el 80%; sin embargo, para cuantificar correctamente la reducción de la movilidad, se debe calcular e incorporar una línea base en el análisis de la movilidad. También es necesaria una línea base para crear correlaciones cuantitativas entre la movilidad reducida y los indicadores COVID19, los cambios económicos y los cambios en las políticas, especialmente a medida que se relajan las restricciones de movilidad.

Conclusión: Aunque la movilidad orgánica varía por provincia, la movilidad bajó de forma general desde los 25 km por día por persona a 5 km por día por persona, a nivel nacional.

**2.** Cada noticia emitida sobre el confinamiento y COVID tiene un efecto medible en la movilidad. El cambio sobre el la línea base puede ayudarnos a entender qué políticas tienen mayores efectos y en cuanto tiempo tardan estas en verse reflejadas en la movilidad a nivel general.

Conclusión: La movilidad se redujo durante 5 días desde el anuncio de estado de emergencia. La restricción de toda actividad no esencial comenzó a afectar significativamente a la movilidad unos 5 días después y, como era de esperar, tuvo el mayor impacto en la movilidad.

**3.** Las políticas de cuarentena tienen un efecto medible en los casos, hospitalizaciones y muertes de COVID.

Conclusión: Los datos indican que las políticas de cuarentena han requerido de 15 días para llevar a cabo la reducción del total de nuevos casos y nuevos fallecimientos.

**4.** El análisis de indicadores que afectan a la evolución del virus o del riesgo en un área geográfica determinada de manera aislada podría dar lugar a conclusiones incorrectas o evaluaciones sesgadas. Los índices multivariables producen un análisis más completo. Según la densidad de población, Madrid y Barcelona representan el mayor riesgo de contagio en España. Sin embargo, la introducción de otras variables como la densidad de la población en riesgo y especialmente en relación con la cobertura de puntos críticos de interés (hospitales, farmacias, supermercados), permite reconocer que provincias requieren de medidas de protección adicionales.

Conclusión:

- Castellón, Cantabria y Guipúzcoa tienen la mayor concentración de poblaciones de riesgo proporcionalmente
- Toledo, Segovia, Salamanca o Navarra tienen menos cobertura hospitalaria respecto a la población vulnerable

**5.** El uso de un baseline<sup>1</sup> para realizar un análisis hiperlocal, u observar el tráfico de pies en el radio inmediato de un PDI en particular, nos ayudará a comprender los patrones de consumo físico desencadenados por la nueva normalidad. Comprender los patrones de consumo ayudará a las empresas y marcas a comunicarse mejor con el nuevo consumidor. Del mismo modo, el monitoreo de los cambios en la navegación digital, no sólo a través de las categorías de contenido, sino también con respecto a otros indicadores como la conexión de red informa los patrones de consumo.

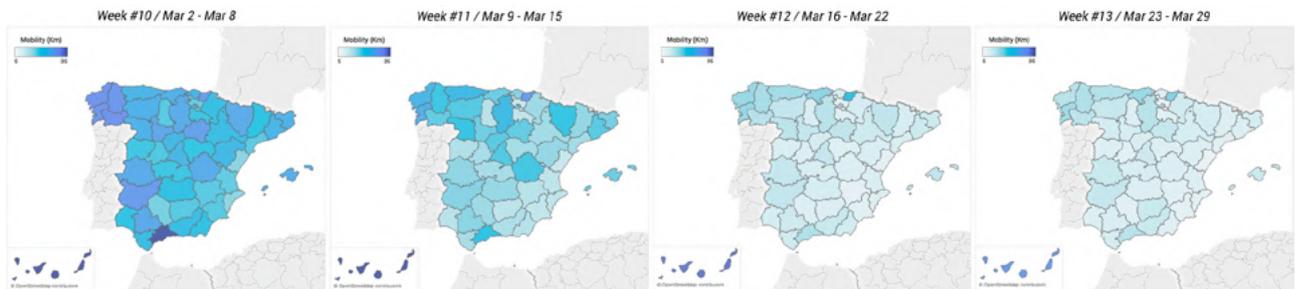
Conclusión: El tráfico de internet se ha disparado un 400%, las categorías con mayor crecimiento son de entretenimiento y compras online. El consumo online en confinamiento es prácticamente igual todos los días de semana aunque el Miércoles y Jueves son los días de menor incidencia.

1. Baseline. El baseline es el valor promedio de la movilidad en los 20 días previos al confinamiento.

# Análisis Geográfico de Datos Dinámicos

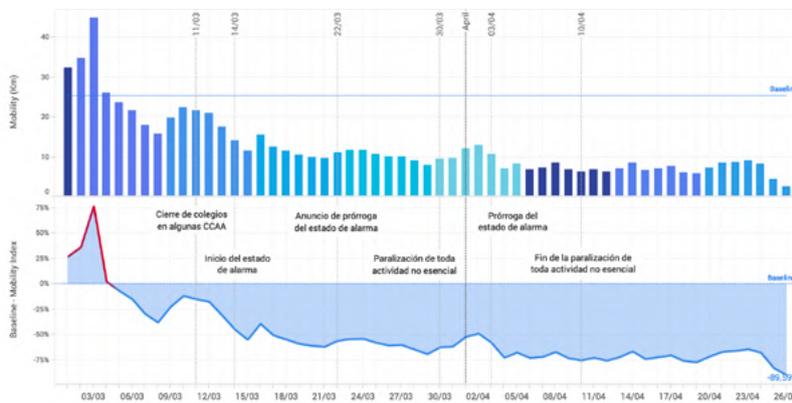
## MOVILIDAD POBLACIONAL

Fig. 1. Variación semanal de movilidad por provincia durante marzo de 2020. Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



Análisis de la movilidad media por usuario y provincia (KM diarios). En origen, hay provincias que ofrecen mayor dispersión que otras, lo que requiere mayor movilidad. Observamos que a medida que evoluciona el confinamiento, la movilidad entre todas las provincias se tiende a homogeneizar y reducir.

Fig. 2. Variación de la movilidad nacional, España. Fuente: Sonata LI (TAPTAP) | UC3M | Predicio

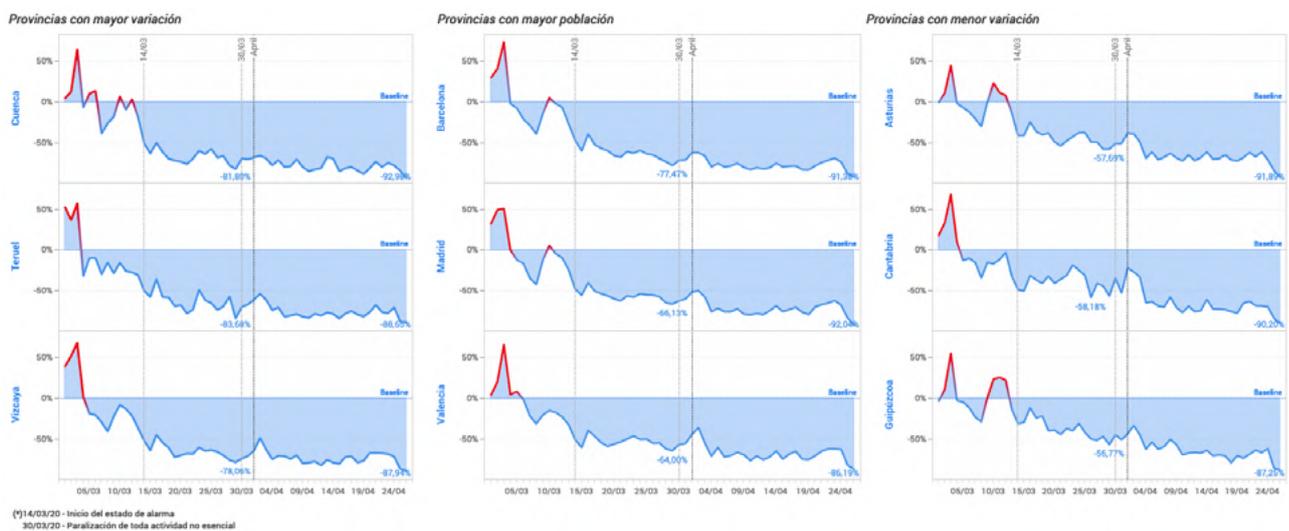


En esta gráfica podemos observar los cambios relativos de movilidad con respecto a un baseline basado en datos pre-COVID19. Dicho baseline supone un dato de referencia que nos sirve para comparar patrones de movimiento entre periodos.

En este caso, el baseline es el valor promedio de la movilidad en los 20 días previos al confinamiento.

4

Fig. 3. Variación de la movilidad nacional. Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



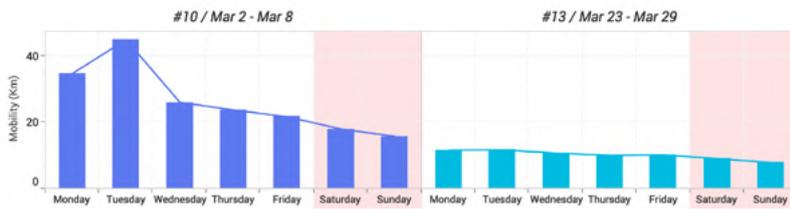
Las tasas de movilidad promedio cambian de 25 km por día a solo 5 kms con respecto al baseline, lo que supone una reducción en la movilidad de hasta el 80%.

Cada noticia emitida sobre el confinamiento significa una disminución en la movilidad semanal aunque no todas las provincias de forma homogénea. Cuenca y Teruel son las provincias con mayor disminución en la movilidad. En el otro extremo, Asturias y Cantabria son las que menor variación en la movilidad registran.

OTROS INDICADORES

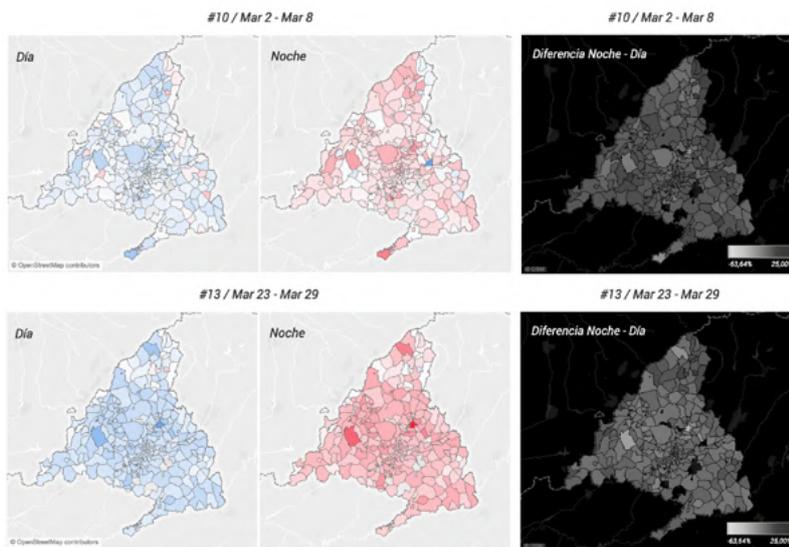
Variación de la movilidad entre los días de la semana y los fines de semana

Fig. 4. Variación de movilidad nacional entre las semanas 10 y 13, España. Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



Durante una semana típica, la movilidad de los usuarios entre semana es aproximadamente el doble que los fines de semana. Sin embargo, en confinamiento, existe una mínima diferencia en movilidad entre semana con respecto a los fines de semana, un 10-15%.

Fig. 5. Cambios en la densidad del tráfico digital por código postal por día entre las semanas 10 y 13. Madrid, España. Fuente: Sonata LI (TAPTAP)



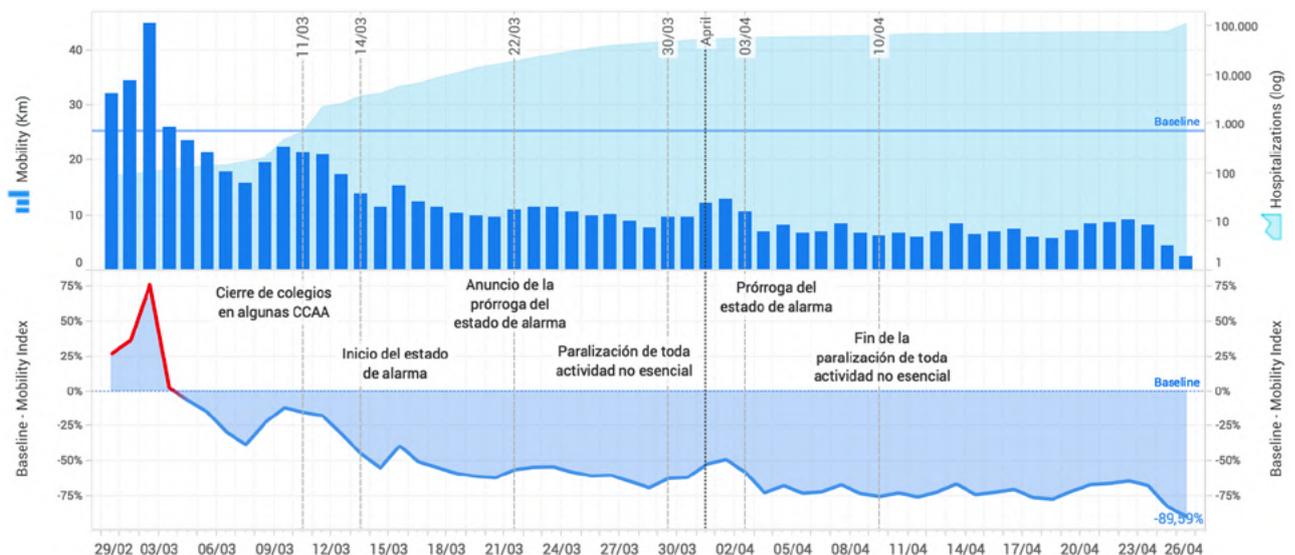
Poniendo como ejemplo la Comunidad de Madrid mostramos las diferencias de densidad del tráfico por código postal entre el día y la noche, derivado de la ubicación de dispositivos móviles.

El gráfico azul muestra la densidad de los usuarios durante el día, y el gráfico rojo, la densidad total por la noche, por código postal.

El gráfico negativo (escala de grises) muestra la diferencia entre el día y la noche, donde las áreas más claras indican menos variación y más, las áreas más oscuras.

La semana 10 muestra una mayor disparidad entre el día y la noche, especialmente en la periferia, que suele ser más comercial que residencial (por lo que esperamos más tráfico durante las horas diurnas). En la semana 13, para cuando el confinamiento era generalizado, la densidad por cada código postal entre el día y la noche es prácticamente la misma.

Fig. 6. Movilidad y hospitalizaciones por COVID-19, España. Fuente: Sonata LI (TAPTAP) | Ministerio de Sanidad | UC3M | Predicio | Tamoco



La tasa de movilidad frente a la tasa de hospitalizaciones a escala logarítmica representa los aumentos exponenciales en los datos sin el sesgo de los resultados de los valores atípicos.

El análisis multivariable determina que la tasa de hospitalizaciones comienza a aplanarse 15 días después del comienzo del inicio del estado de alarma, por lo que se tarda aproximadamente este tiempo en que los beneficios del confinamiento surtan efecto.

Existen correlaciones claras entre las políticas de reclusión y la movilidad y por lo tanto, las hospitalizaciones. Cada noticia emitida sobre el confinamiento le sigue una disminución en la movilidad semanal.

## Análisis Geo y Población

### Análisis Demográfico de Vulnerabilidad

Los siguientes cuatro gráficos muestran la distribución de la población por volumen total, y luego por grupo de edad.

Fig. 7. Distribución de la población total, por provincia.



Fig. 8. Distribución de la población vulnerable, por provincia.



Fig. 9. Distribución de la población joven, por provincia

Fuente: Sonata LI (TAPTAP) | INE



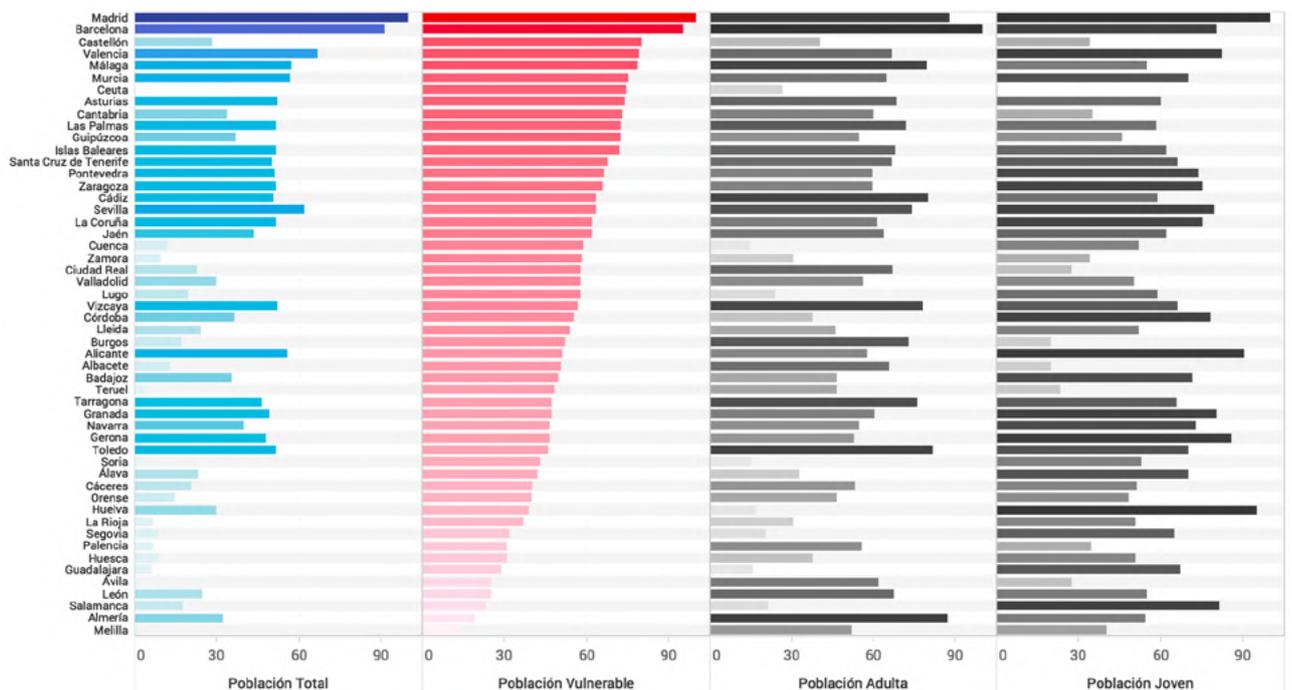
Fig. 10. Distribución de la población adulta, por provincia.



6

En fig 8, si clasificamos cada provincia por su densidad de población vulnerable, vemos que Castellón, Cantabria o Guipúzcoa tienen un riesgo más alto, incluso si los niveles de población total son más bajos.

Fig. 11. Índice de tipo de población por grupo de edad y volumen total. Fuente: Sonata LI (TAPTAP) | INE



El uso de índices multivariables nos permite analizar los datos y las provincias contra más de un factor, lo que muestra una imagen más completa. La figura 11 muestra los datos que incluyen tanto el volumen total de población vulnerable, como también la proporción (%) de esta con respecto al total. Madrid, Barcelona y Castellón siguen siendo provincias con riesgo para la población vulnerable, pero si atendemos al índice multivariante, Valencia y Málaga se colocan en la parte superior de la lista.

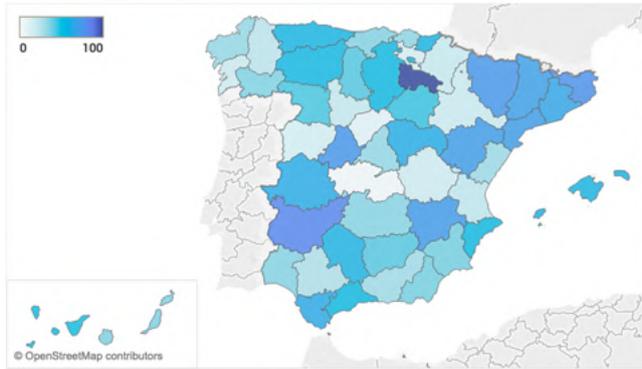
### ANÁLISIS CONTEXTUAL: PUNTOS DE INTERÉS COBERTURA

Incorporando datos de movilidad sobre la densidad de diferentes puntos de interés (PDIs) nos ayuda a comprender que las áreas con menos PDIs requieren más movilidad y que las áreas con una mayor saturación de usuarios por PDI pueden generar un mayor riesgo de exposición al virus.

#### Cobertura de servicios hospitalarios y de salud por densidad de población

Fig. 12. Índice de cobertura hospitalaria por densidad de población total.

Fuente: Sonata LI (TAPTAP) | INE



Por ejemplo, el número de hospitales por región no es muy significativo analizado de forma aislada. Si bien en Fig. 13 observamos la cobertura hospitalaria respecto a cada grupo de edad. La Rioja, Teruel y Albacete tienen una buena cobertura hospitalaria por densidad de población, sin embargo su índice de vulnerabilidad demográfica no es alto, pudiendo suponer esto una ventaja frente a contagios. Son resaltables las coberturas tan bajas a nivel hospitalario de Toledo, Segovia, Salamanca o Navarra, pudiendo esto ser un problema a la hora de atender casos de riesgo.

(No se han tenido en cuenta para los cálculos ni el número de camas hospitalarias ni los hospitales improvisados en centros como IFEMA)

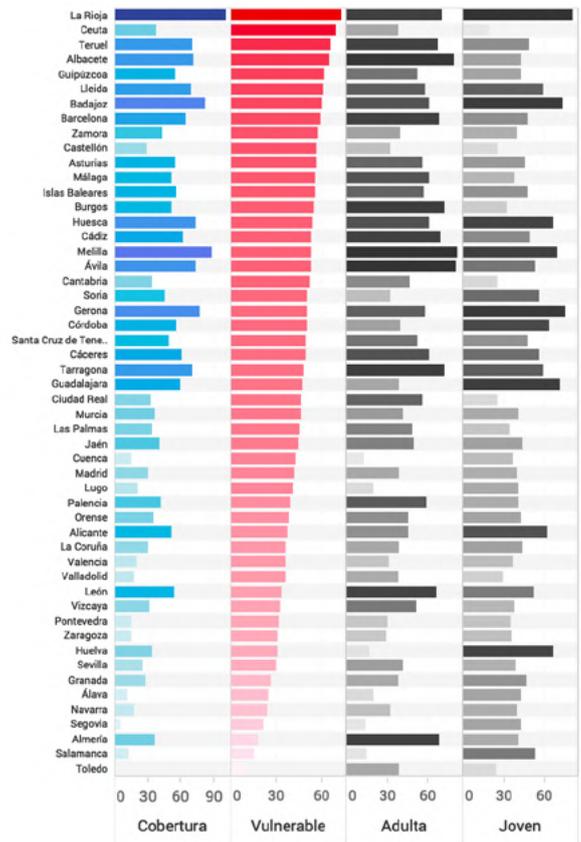
Fig. 14. Análisis de PDI hiperlocal: tráfico en el Hospital Severo Ochoa de Leganés (Madrid, España). Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



El análisis a nivel Geo puede ser estudiado de forma agregada sobre el territorio, pero también detallado a nivel punto de interés.

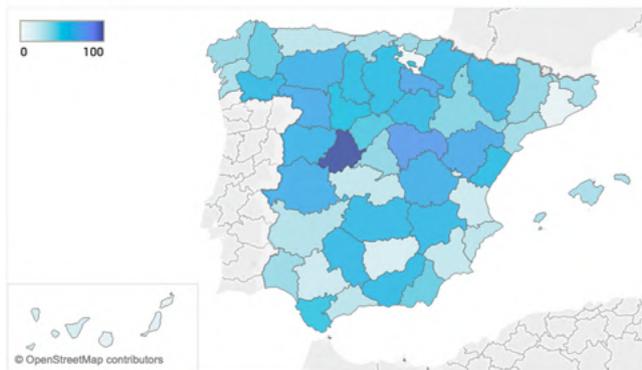
En el gráfico, datos individualizados por punto (Hospital Severo Ochoa de Leganés, Madrid) permiten analizar de forma individualizada el tráfico diario de usuarios en la zona y compararlo con su media de referencia o Baseline.

Fig. 13. Índice de cobertura hospitalaria por volumen de población y grupo de edad. Fuente: Sonata LI (TAPTAP) | INE



### Cobertura de farmacias por densidad de población

**Fig. 15. Índice de cobertura de farmacias por densidad de población total**  
Fuente: Sonata LI (TAPTAP) | INE



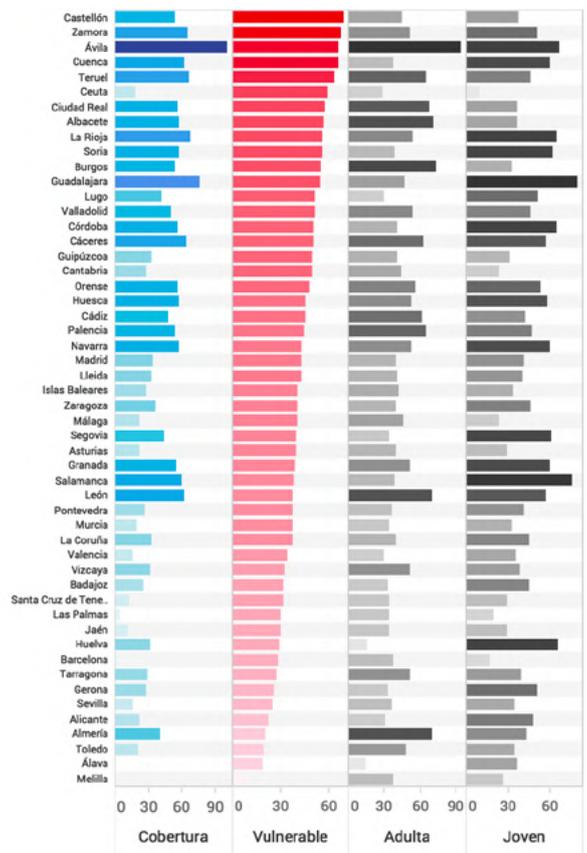
Se muestra como Barcelona y Madrid tienen menos farmacias por persona aunque lideran en densidad de población, al revés que Ávila, Zamora y Castellón que tienen mayores tasas de cobertura farmacéutica con respecto al índice de población vulnerable, pudiendo atender mejor la demanda que la enfermedad pudiera requerir.

**Fig. 17. Análisis de PDI hiperlocal: tráfico en una farmacia residencial en Madrid, España.** Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



El análisis hiperlocal de la farmacia muestra que aunque el tráfico peatonal cae por debajo del baseline después de comenzar el estado de alarma, experimenta aumentos y disminuciones constantes en el tráfico y no una caída continua. Esto se debe a que las farmacias son un servicio fundamental que se ocupa de la atención sanitaria.

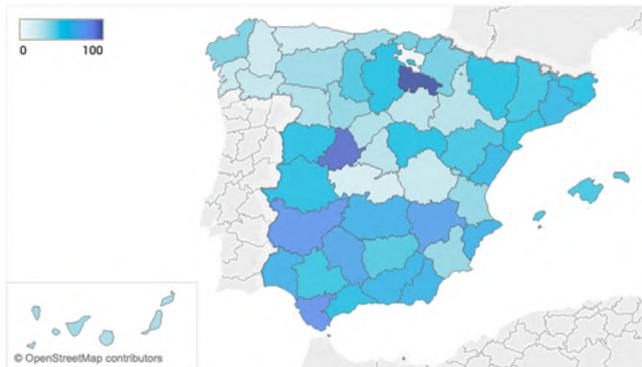
**Fig. 16. Índice de cobertura de farmacias por volumen de población y grupo de edad.** Fuente: Sonata LI (TAPTAP) | INE



### Cobertura de supermercados y servicios de alimentación por densidad de población

**Fig. 18. Índice de cobertura de supermercados por densidad de población**

Fuente: Sonata LI (TAPTAP) | INE



El gráfico muestra un análisis hiperlocal que proporciona el tráfico de clientes a un supermercado en Alcobendas. Aunque las tiendas de alimentación son PDIs muy visitados, el tráfico ha caído en más del 90% con respecto a su baseline. La reactivación económica por fases nos permitirá analizar cómo varía el tráfico hiperlocal, regional y nacional y cómo evoluciona la pandemia respectivamente.

**Fig. 20. Análisis de PDI hiperlocal: tráfico en un supermercado Alcobendas (Madrid, España).**

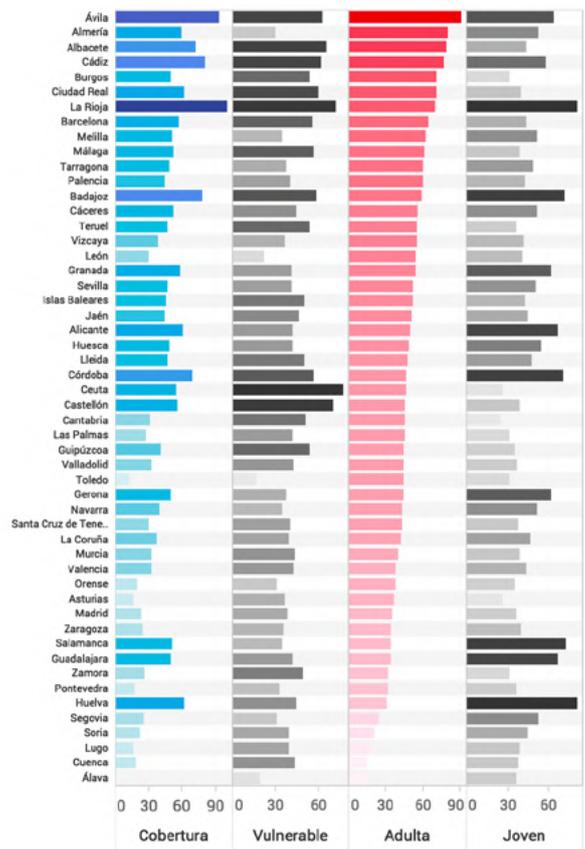
Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



El gráfico muestra un análisis hiperlocal que proporciona el tráfico de clientes a un supermercado en Alcobendas. Aunque las tiendas de alimentación son PDIs muy visitados, el tráfico ha caído en más del 90% con respecto a su baseline. La reactivación económica por fases nos permitirá analizar cómo varía el tráfico hiperlocal, regional y nacional y cómo evoluciona la pandemia respectivamente.

**Fig. 19. Índice de cobertura de supermercados por densidad de población y grupos de edad.**

Fuente: Sonata LI (TAPTAP) | INE



## Cobertura de estaciones y servicios de transporte por densidad de población

Fig. 21. Índice de cobertura de estaciones de transporte por densidad de población. Fuente: Sonata LI (TAPTAP) | INE

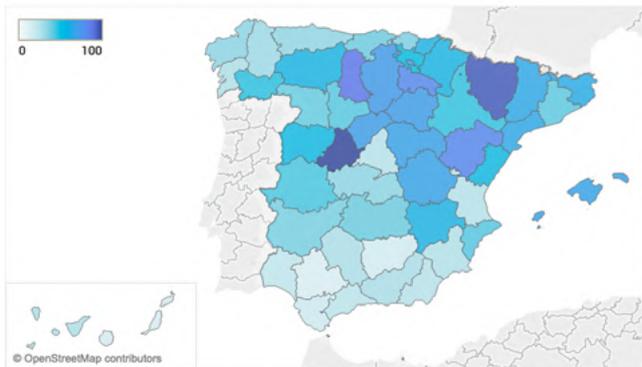


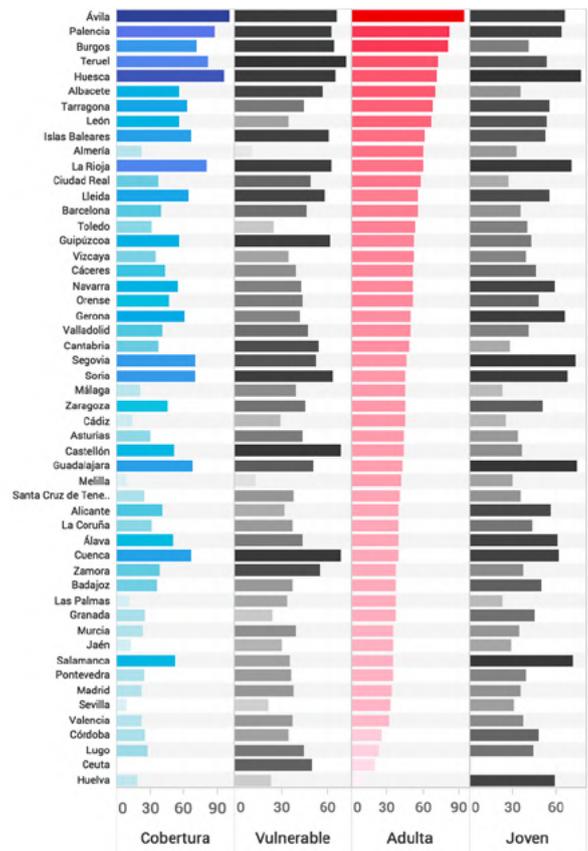
Fig. 22 muestra que Castilla y León tienen la tasa de cobertura de servicios de transporte más alta por población adulta, el segmento con mayor probabilidad de utilizar estos servicios. Madrid y Barcelona, pese a tener menor cobertura pueden presentar mayores necesidades de estos servicios debido a la alta densidad de población y elevar el riesgo de contagio debido a este uso.

Fig. 23. Análisis de PDI hiperlocal: tráfico en la Estación de Atocha (Madrid, España). Fuente: Sonata LI (TAPTAP) | UC3M | Predicio



El gráfico muestra un análisis hiperlocal que proporciona el tráfico a pie de clientes a una de las estaciones más concurridas del país. La estación de Atocha en Madrid donde el tráfico ha caído en más del 90% con respecto a su baseline. La reactivación económica por fases nos permitirá analizar cómo varía el tráfico hiperlocal, regional y nacional y cómo evoluciona la pandemia respectivamente.

Fig. 22. Índice de cobertura de estaciones de transporte por volumen de población y grupo de edad. Fuente: Sonata LI (TAPTAP) | INE



## Análisis Geo Multivariable

### ANÁLISIS NACIONAL COVID

Fig. 24. Evolución de casos, defunciones, hospitalizaciones, pacientes recuperados y UCI. Datos acumulados. Fuente: Ministerio de Sanidad

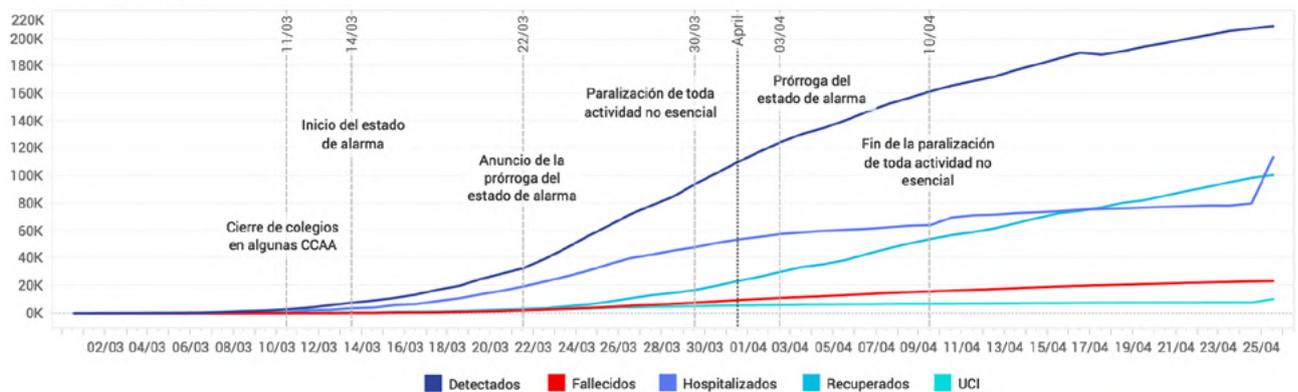
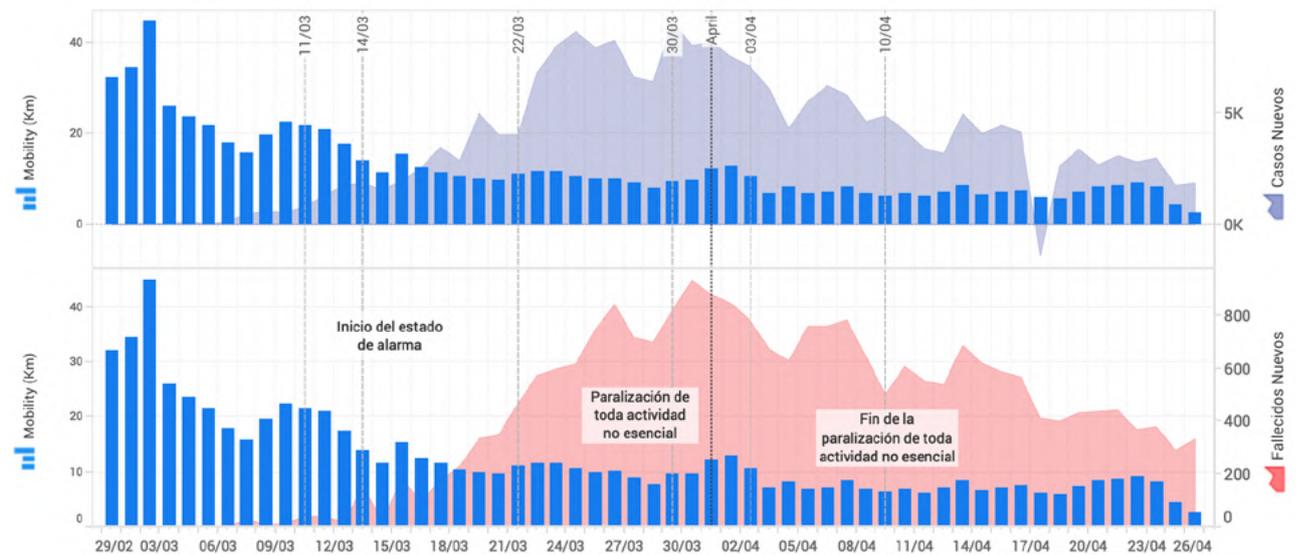


Fig. 25. Movilidad e indicadores COVID-19. Fuente: Sonata LI (TAPTAP) | Ministerio de Sanidad | UC3M | Predicio | Tamoco



11

Tanto los fallecimientos como los casos nuevos alcanzan su máximo alrededor del 1 de abril —unas 2 semanas después de la activación de las medidas de cuarentena— y luego comienzan a decrecer. Como los nuevos casos siempre preceden a los fallecimientos, la tasa de nuevos casos se ralentiza más rápido que la tasa de muertes, pero ambos siguen la misma tendencia a la baja.

Aunque el ritmo de nuevos fallecidos es consecuencia no solo de la movilidad, el pico de fallecidos aparece aproximadamente a los 15 días desde que la movilidad está contenida de forma estable.

### Casos acumulados y fallecimientos por CCAA

Fig. 26. Tasa de incidencia acumulada total. Fuente: Ministerio de Sanidad

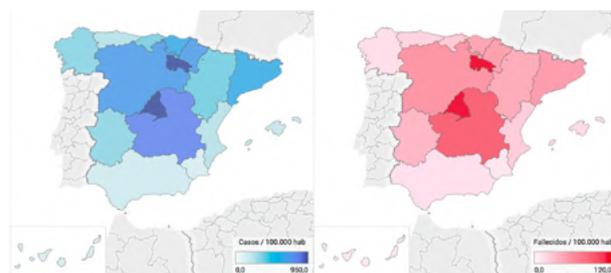
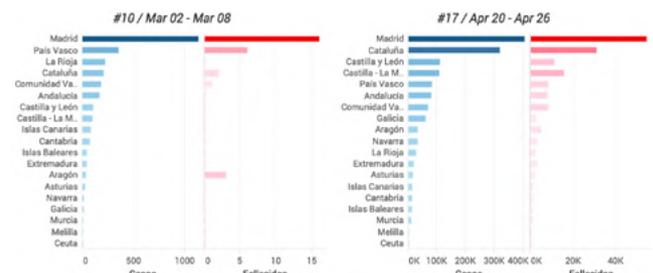


Fig. 27. Casos acumulados y muertes por CCAA. Fuente: Ministerio de Sanidad

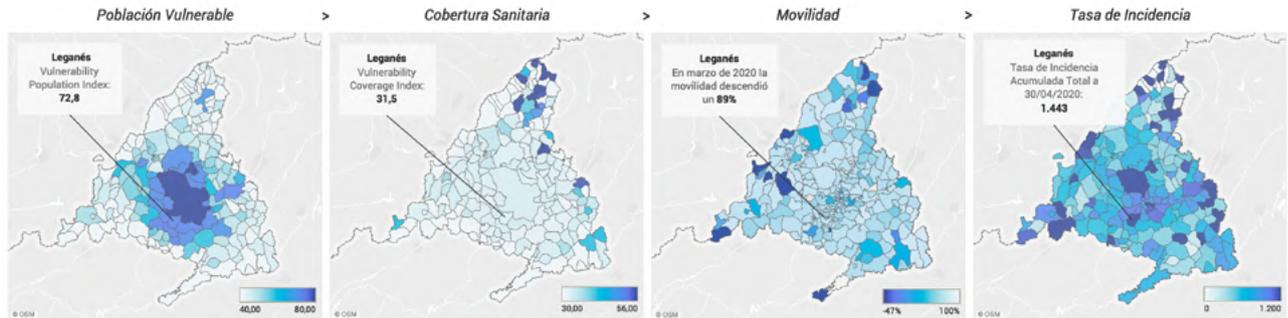


Podemos monitorizar la evolución de la pandemia a nivel geográfico atendiendo a variables contextuales, de población o de movilidad para comprender más sobre cómo se desarrolla y establecer relaciones entre los datos.

## ANÁLISIS REGIONAL COVID

En la siguiente secuencia, se muestran los datos individuales de los 3 bloques de análisis estudiados: Demografía (vulnerabilidad medida por edad), Contexto Geo (Cobertura de Hospitales por localidad), Comportamiento (Movilidad de los usuarios). Tomando como referencia una población como Leganés, con una de las mayores incidencias de COVID19 en la comunidad de Madrid (Mapa 4), observamos una muy alta correlación entre los 3 grupos de variables anteriormente mencionados.

Fig. 28. Correlación entre indicadores, Leganés (Madrid) Fuente: Sonata LI (TAPTAP) | INE

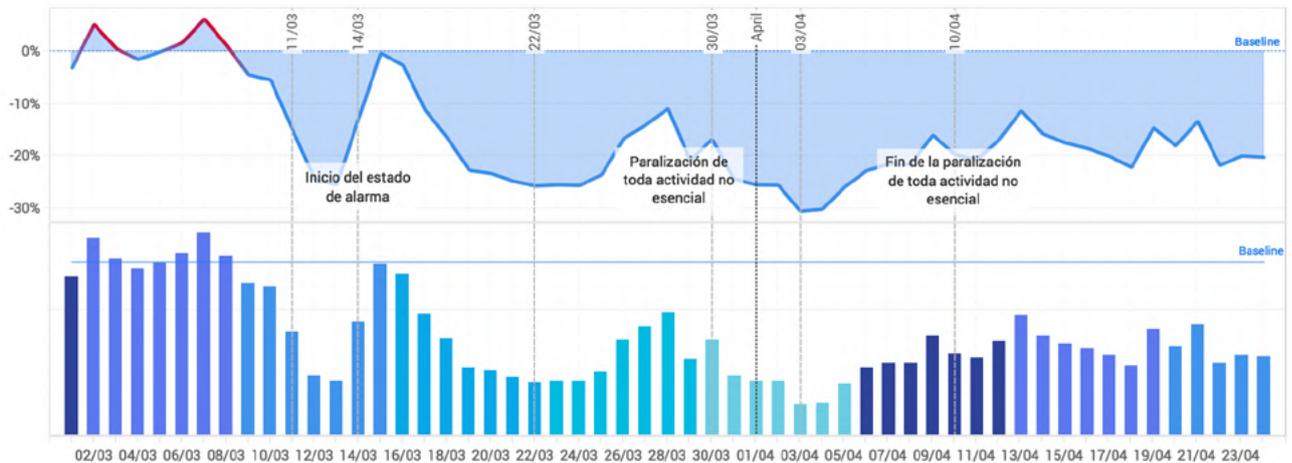


En consecuencia, estos conjuntos de datos dan la posibilidad de construir un índice multivariable a nivel GEO que explique y prevea el posible grado de afectación o incidencia por territorios.

## Audiencias. Cambios del comportamiento

### COMPORTAMIENTO DIGITAL

Fig. 29. Variación del tráfico digital con conexión 3G/4G. Fuente: Sonata LI (TAPTAP)



El Baseline en este gráfico representa el promedio de dispositivos que usan redes móviles para conectarse a Internet antes de confinamiento. Las conexiones a la red móvil se redujeron en cerca de un 30% en relación con el Baseline, mientras que las de wifi experimentaron un aumento equiparable. Este cambio se debe a una disminución de la movilidad ya que se trata de conexiones de ubicaciones estáticas como casa o el trabajo.

Después del 12 de abril, las conexiones de red experimentaron una subida, lo que indica una mayor movilidad después del período de bloqueo inicial.

13

Fig. 30. Compras en Centro Ciudad (No confinamiento). Fuente: Sonata LI (TAPTAP)

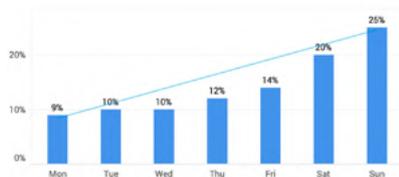


Fig. 31. Compras en CC Periferia (No confinamiento). Fuente: Sonata LI (TAPTAP)

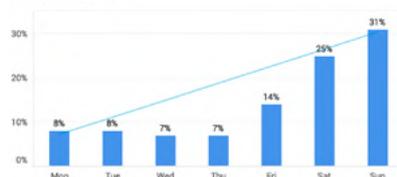
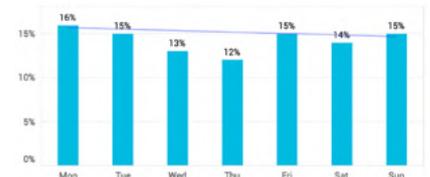


Fig. 32. Compras en confinamiento. Fuente: Sonata LI (TAPTAP)



#### Comportamiento concentrado pre confinamiento

En periodos previos, existe un **claro patrón de comportamiento en el punto de venta físico**. Concentración alrededor del fin de semana, más acusado en centros comerciales que en centros urbanos.

*Casi la mitad de las compras físicas se concentran en tan sólo dos días de la semana.*

#### Desestacionalización de compras

Los hábitos individuales cambian y los hábitos de compra física también. Los días más bajos pasan a ser los picos de visita, mientras que los fines de semana se equilibran o quedan incluso por debajo de la media diaria.

*Linealidad en el hábito de compra, completamente desvinculada de la compra / ocio.*

## APÉNDICE:

### Trabajando con Location: Una perspectiva técnica

Los datos de localización mencionados anteriormente son recolectados de forma anónima a través de aplicaciones móviles que hayan permitido el acceso a dicha información. En cada dispositivo móvil, los sistemas operativos más populares (iOS & Android) combinan multitud de tecnologías de geolocalización como el GPS, Wifi o Bluetooth para poder proporcionar a los usuarios una geolocalización precisa. Siendo la más precisa y la usada para este estudio, el GPS, aquella que es capaz de identificar al dispositivo en un punto (latitud, longitud) determinado dentro de un mapa o sistema de referencia. Para facilitar y escalar los análisis de datos de localización, estos puntos (lat,long) son transformados en en geohash, un sistema de codificación universal utilizado para identificar localizaciones según diferentes niveles de precisión y tamaño. Las coordenadas (latitud & longitud) son hasheadas en áreas o bloques que concentran muchos puntos o coordenadas acorde al punto de interés, o en este caso, el objetivo del análisis. Además de servir como metodología de análisis, transformar estas coordenadas en geohashes conlleva un beneficio adicional de cara a proteger la privacidad de los usuarios. Esencialmente, mediante este proceso hacemos un zoom hacia fuera desde un punto concreto y lo representamos en un área más o menos grande en función del nivel de geohash. De esta forma, inyectamos lo que denominamos "ruido" a la localización al meterle un factor aleatorio de error controlado. Para llevar a cabo este estudio se ha elegido en función de cada caso, un nivel de geohash que, además de preservar la privacidad del usuario, asegura un nivel de precisión necesario para el análisis de datos, pudiendo ser áreas que cubren decenas de metros, o cientos de metros dependiendo del objetivo de cada análisis. Para medir la movilidad de los usuarios, se ha utilizado una técnica conocida como "Radius of Gyration" que mide cuánto se ha movido cada usuario desde su centro de trayectoria, que es el sitio donde más veces pasa un usuario, e.g, su lugar de residencia. Con esta técnica podemos calcular cuantos Km se mueve en media un usuario al día.

### Una Muestra Representativa

Para garantizar que la precisión y cobertura del estudio, se ha trabajado y analizado una muestra de datos de dispositivos y localizaciones que representan de forma estadística la población española. Para poder configurar dicha muestra de datos, TAPTAP y sus colaboradores han llevado a cabo un proceso de limpieza y filtrado de millones de dispositivos que generan cientos de millones de datos de localización en un periodo de tiempo. Para poder llevar a cabo un análisis de movilidad riguroso y preciso, es necesario utilizar solo aquellos dispositivos que han generado multitud de señales de localización de forma continuada y distribuida a lo largo del día (algo importante para poder justificar cambios sustanciales de movilidad y no micromovimientos a nivel local, como sacar la basura). Más concretamente, para éste análisis se requieren al menos 5 muestras de datos de localización por dispositivo y por día separados al menos 8 horas. Estos datos de localización son generados de forma diferente a lo largo del día dependiendo de hábitos de consumo digital o de las características/configuración de cada aplicación, siendo la frecuencia mayor o menor en cada caso. En este caso, es clave contar con colaboradores como Predicio o Tamoco que nos proporcionan datos de gran precisión y alta frecuencia de muestreo (alrededor de 100-200 muestras diarias por dispositivo/usuario). El resultado final de la agrupación de estas señales bajo dichas características, que incluyen ID anónimo y hasheado del dispositivo, tiempo y fecha, localización, y otros metadatos no-personales es lo que definimos como un (Evento de Localización). Este objeto de localización es la unidad individual, compuesta por todos estos parámetros, que nos permite analizar en profundidad el movimiento y el contexto de para cada evento generado.



EUROPE | AFRICA | NORTH AMERICA | SOUTH AMERICA

