familias, igualdad y bienestar social | MADRID

Dirección General de Innovación
y Estrategia Social

Estudio de análisis georreferenciado predictivo y elaboración de cuadros de mando ante el impacto socioeconómico del COVID-19 sobre la población vulnerable y demanda asistencial de la ciudad de Madrid

Noviembre 2020

esri España
THE SCIENCE OF WHERE

FICHA TÉCNICA

FICHA TÉCNICA DEL ESTUDIO	
Título	Estudio de análisis georreferenciado predictivo y elaboración de cuadros de mando ante el impacto socioeconómico del COVID-19 sobre la población vulnerable y demanda asistencial de la ciudad de Madrid
Fichero de partida	Fichero de llamadas geolocalizadas de solicitudes de asistencia social, a través del canal telefónico 010 de Línea Madrid, realizadas desde el inicio del estado de alarma por covid 19 hasta el 3 de junio de 2020.
Otros ficheros de carácter interno	Tramitaciones iniciadas registradas en el sistema de gestión de los servicios sociales del Ayuntamiento de Madrid -CIVIS-.
Otras fuentes de datos externas	Datos de Padrón y seccionado censal de la ciudad de Madrid Paro, renta, afiliación, nivel educativo y tipología de hogar de la ciudad de Madrid
Elaboración del informe	Esri España Soluciones Geoespaciales S.L.
Adaptación del informe	Dirección General de Innovación y Estrategia Social
Periodo de ejecución	Del 11 de junio al 18 de septiembre de 2020
Número de Expediente	171/2020/00484 Presupuesto base de licitación: 14.999,00 € (sin Iva) Importe IVA 21%: 3.149,79 € (Iva) Presupuesto total (IVA incluido 21%): 18.148,79 € (con Iva)
Empresa adjudicataria y precio de adjudicación	Esri España Soluciones Geoespaciales S.L. Precio de adjudicación: 11.950,00 € (sin Iva) 2.509,50 € (Iva) 14.459,50 € (con Iva)

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

ÍNDICE

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

1. Introducción

1.1. Resumen ejecutivo

El Ayuntamiento de Madrid tiene atribuida, a través de la Dirección General de Innovación y Estrategia Social (DGIyES), la competencia de facilitar al conjunto de la organización un conocimiento global, permanente y actualizado de las necesidades y aspiraciones sociales de la ciudadanía madrileña y específicamente de los grupos y poblaciones más vulnerables. Por ello, desde el inicio de la pandemia de COVID-19 se vio la necesidad de analizar su efecto en la demanda y tramitación de esta asistencia.

Tras la declaración del estado de alarma por pandemia covid19 (Real Decreto 463/2020, de 14 de marzo) los centros de servicios sociales del Ayuntamiento de Madrid permanecieron, en su mayoría, físicamente cerrados desde el 16 de marzo hasta el 3 de junio de 2020, aunque su funcionamiento siguió de forma online. Durante ese periodo de tiempo las solicitudes de asistencia social se redirigieron al canal telefónico 010 de Línea Madrid.

Una vez recibida la solicitud, desde los servicios sociales se procedió al inicio de la tramitación correspondiente en el caso de que la solicitud se estimara procedente.

Hay que señalar que del total de llamadas recibidas en el 010, 114.353 llamadas de 74.887 personas, sólo se pudieron geolocalizar 86.287 llamadas que hicieron 54.817 personas. El presente estudio se ha elaborado a partir del fichero de llamadas geolocalizadas, de la información registrada en el sistema de gestión de servicios sociales- CIVIS- del Ayuntamiento de Madrid , así como de otras fuentes externas (datos de paro, Padrón, afiliación.. de la ciudad de Madrid).

Conviene remarcar que los datos y las conclusiones de este estudio corresponden a un determinado periodo de una situación coyuntural extraordinaria. El análisis de un periodo más amplio permitirá obtener conclusiones más fidedignas de la realidad.

El análisis se ha realizado con ArcGIS, herramienta GIS corporativa del Ayuntamiento de Madrid.

A continuación se presenta el alcance, la descripción de procesos, el análisis y las principales conclusiones del estudio.

1.2. Alcance

Entre los principales objetivos del estudio cabe destacar:

a) Contribuir al dimensionamiento de los efectos socioeconómicos por pandemia Covid19 sobre la población de Madrid, geolocalizándolos y realizando predicciones de su evolución a corto y medio plazo mediante técnicas analíticas avanzadas (elaboración de mapas de calor…)

b) Analizar los perfiles de las nuevas demandas de asistencia por parte de los servicios sociales comparándolos con los de las personas usuarias registradas con anterioridad a la crisis por Covid 19.

c) Modelizar la demanda asistencial en función de parámetros sociodemográficos y socioeconómicos del territorio.

1.3. Datos utilizados en el estudio

A continuación se relacionan los ficheros de datos utilizados en el estudio:

- Padrón (INE).
- Seccionado censal (INE).
- Renta de 2019 (Michael Bauer). Dato procedente de Esri Demographics, derivado de datos de INE (metodología disponible aquí).

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

- Datos de nivel educativo y tipologías de hogar (AIS). Dato procedente de Esri Demographics, (información disponible aquí).

- Paro por sección (ciudad de Madrid): Comparativa de la serie anterior –serie actual del paro registrado, desde mayo de 2005 a diciembre de 2013, desagregado por sexo. Paro por sección censal desagregado por sexo, grupo de edad y sector de actividad en 2006, 2007, 2008, 2017, 2018, 2019 y junio 2020. INE. Elaboración Ayuntamiento de Madrid. Subdirección General de Estadística.

- Afiliación: Personas afiliadas que residen en la ciudad de Madrid por sección censal (desagregado por sexo). Años 2007, 2008, 2018, 2019 y 2020. INE. Elaboración Ayuntamiento de Madrid. Subdirección General de Estadística.

- Llamadas al teléfono 010 durante el periodo del 19 de marzo al 3 de junio 2020.Ayuntamiento de Madrid.

- Tramitaciones iniciadas registradas en el sistema de gestión de los servicios sociales del Ayuntamiento de Madrid -CIVIS- Las tramitaciones con anterioridad al estado de alarma se refieren al periodo del 1 de enero de 2019 al 18 de marzo de 2020; las tramitaciones iniciadas después del estado de alarma asociadas a llamadas geolocalizadas al 010 se refieren al periodo del 19 de marzo al 9 de septiembre de 2020 (última fecha de extracción del fichero). Ayuntamiento de Madrid.

- Fichero actualizado de secciones censales vinculadas a cada centro de servicios sociales (CSS) del Ayuntamiento de Madrid. Distribución del número de trabajadores/as sociales por CSS.

2. Procesado de datos

En esta sección se describen los procesos realizados para convertir los datos de partida en conjuntos de datos geográficos (capas) de manera que puedan utilizarse en mapas y análisis posteriores.

2.1. Datos de demanda y tramitaciones iniciadas

En este proceso se han generado diferentes capas y mapas que permiten ver cómo se comporta la demanda y cómo ha evolucionado antes y después del estado de alarma. Para ello se han usado las distintas herramientas que ofrece la solución ARCGIS Pro.

Uno de los objetivos del proyecto es analizar las tramitaciones antes y después del estado de alarma y determinar si ha cambiado el perfil de la persona demandante. Para ello, se ha comparado el perfil de las personas asociadas a tramitaciones iniciadas desde el 19 de marzo (considerada como fecha de corte) al 9 de septiembre de 2020 con el de las personas asociadas a las tramitaciones realizadas con anterioridad al estado de alarma.

Se parte de los siguientes ficheros:

- Llamadas al 010
- Tramitaciones 2019_2020

En el fichero de tramitaciones se ha detectado que, en algunos registros, el número de menores es mayor de 10. Cuando no es posible identificar el domicilio del solicitante se utilizan determinadas direcciones postales (direcciones de centros, residencias u otros) para registrarlas. Aunque estos registros correspondan a personas reales, la ubicación geográfica puede no ser correcta, contribuyendo a la distorsión de determinados análisis. Por tanto, se han eliminado estos registros anómalos (42 registros) para evitar incluir sesgos en los análisis posteriores.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

A continuación se han convertido los ficheros Excel originales a formato GIS usando la herramienta de exportación Excel to table. Para ubicarlos espacialmente se dispone de un campo en el fichero que indica la sección censal. Para cada sección censal se han agregado los datos obtenidos de cada uno de estos ficheros. Para ello, se ha empleado la herramienta summary statistics, que permite agrupar por un determinado campo y agregar estadísticas de otros campos usando distintas funciones (media, suma, recuento, etcétera).

Se ha ejecutado este proceso para cada uno los ficheros de partida, utilizando el campo de seccionado censal, añadiendo como campos nuevos la media del número de menores en la unidad familiar y la edad media de las personas usuarias.

Este proceso genera una capa de salida con los valores estadísticos y con un campo denominado Frequency, que indica el número de tramitaciones o llamadas (dependiendo del fichero de entrada en cada caso) para cada sección censal.

Una vez se dispone de una capa agregada a nivel de sección censal, se asocia a la geometría del seccionado censal mediante un proceso llamado join (una unión de tablas entre la tabla alfanumérica y la tabla geográfica que contiene las geometrías).

Existen otros dos campos de tipo alfanumérico (categorías) que se han utilizado: familia monoparental y sexo. Para poder añadir estos datos se han seleccionado, en la tabla de tramitaciones, los registros en los que la demandante es MUJER; la herramienta frequency cuenta el número ocurrencias y lo vuelca sobre la capa de secciones censales.

Lo mismo se realiza con aquellos trámites en los que la familia es monoparental. De este modo, se han obtenido campos, para cada sección censal, que expresan el número acumulado de solicitudes asociadas a mujeres y a familias monoparentales.

Hay otras variables sociodemográficas y socioeconómicas que se han incorporado para enriquecer el análisis. En particular, la renta per cápita, que el INE pone a disposición del público para los años 2015, 2016 y 2017. En el catálogo de Esri Demographics se dispone de datos de 2019, por lo que se han incorporado usando la herramienta geoenriquecimiento para añadir las variables a la capa maestra.
Las variables añadidas han sido:

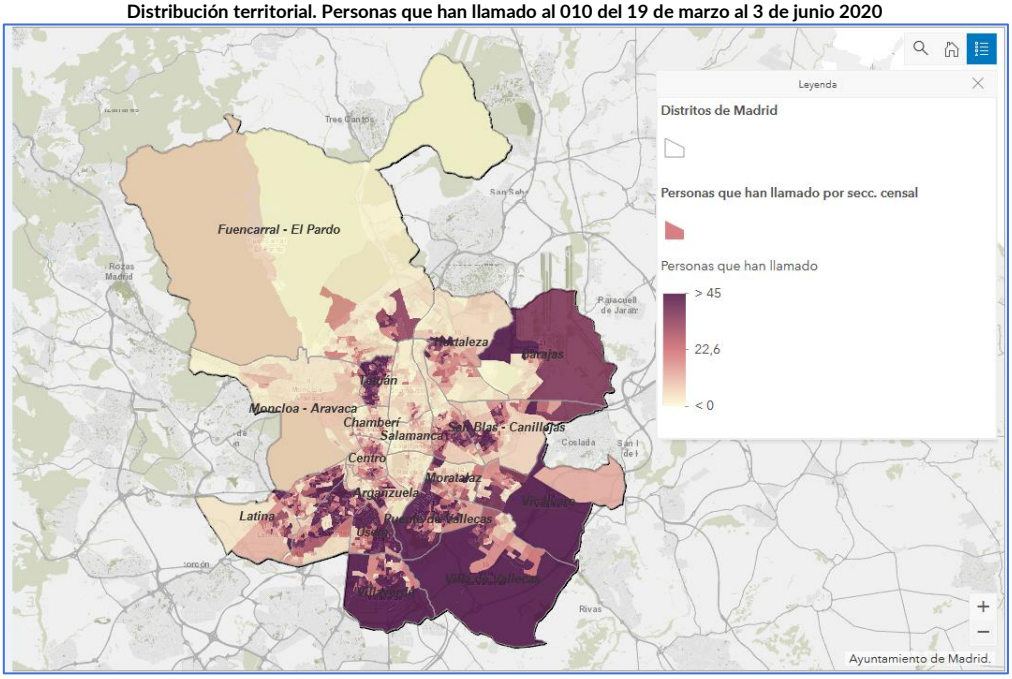
- Renta Per Cápita año 2019
- Educación Secundaria
- Educación Universitaria
- Educación Analfabeto
- Sin Educación
- Educación Primaria
- Estado Civil Soltero
- Estado Civil Casado
- Estado Civil Divorciado
- Tamaño del Hogar
- Número total de Hogares
- Tipología H (Con niños clase social baja)
- Tipología F (Con adolescentes y niños clase social baja)
- Tipología G (Con/Sin jóvenes clase social baja)
- Tipología P (pensionistas en pareja nivel bajo de estudios)
- Tipología Q (pensionistas solitarios nivel bajo de estudios)
- Tipología I (DINK –*double income, no kids*– Clase social baja)

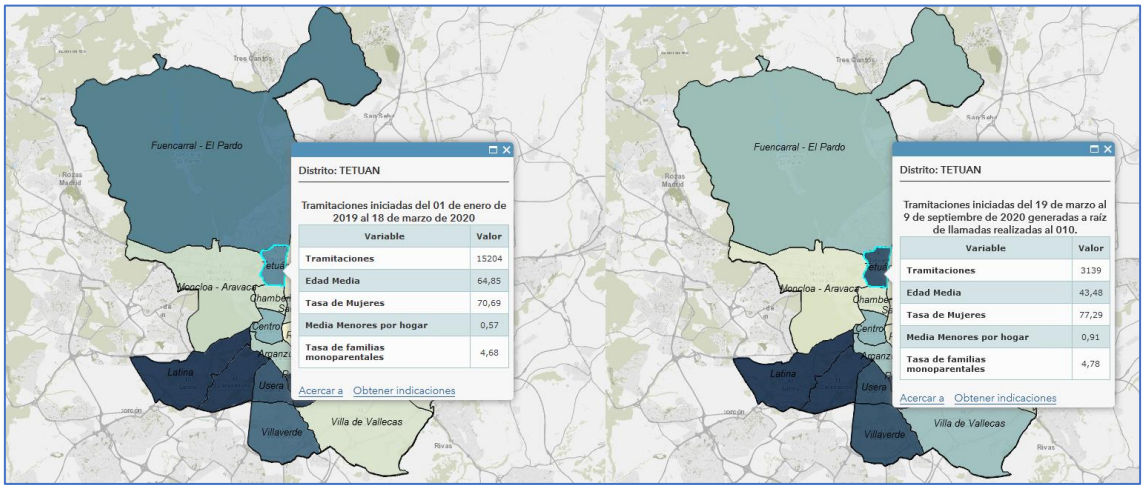
Para añadir las variables se ha usado la herramienta Enrich.

Tras estos procesos se ha obtenido una capa de seccionado censal con información agregada y variables de contexto que va a ser la capa maestra a partir de la cual se han creado diferentes mapas temáticos y análisis descriptivos y explicativos.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Distribución territorial. Personas que han llamado al 010 del 19 de marzo al 3 de junio 2020





Distribución territorial. Tramitaciones iniciadas antes (mapa de la izquierda) y después (mapa de la derecha) del estado de alarma en el distrito de Tetuán

2.2. Clustering

El *Clustering* es un proceso que consiste en agrupar un conjunto de objetos en subconjuntos llamados *Clústers*, donde cada clúster agrupa objetos de características similares entre sí y distintas a las presentes en otros clústers.

Se ha utilizado la herramienta de ArcGIS Multivariate clustering, que agrupa secciones censales a partir de atributos numéricos, de modo que todas las entidades dentro de cada clúster sean lo más parecidas posible, y todos los clúster en sí sean tan diferentes como sea posible. La similitud de las entidades se basa en el conjunto de atributos o variables que sean de interés para el propósito del estudio.

Se parte de la capa maestra de seccionado censal, los datos de llamadas al 010 y las tramitaciones iniciadas (antes y después del estado de alarma), además de las variables de segmentación de población asociadas a ellas: edad media, número de menores en el hogar, número de familias monoparentales y número de mujeres.

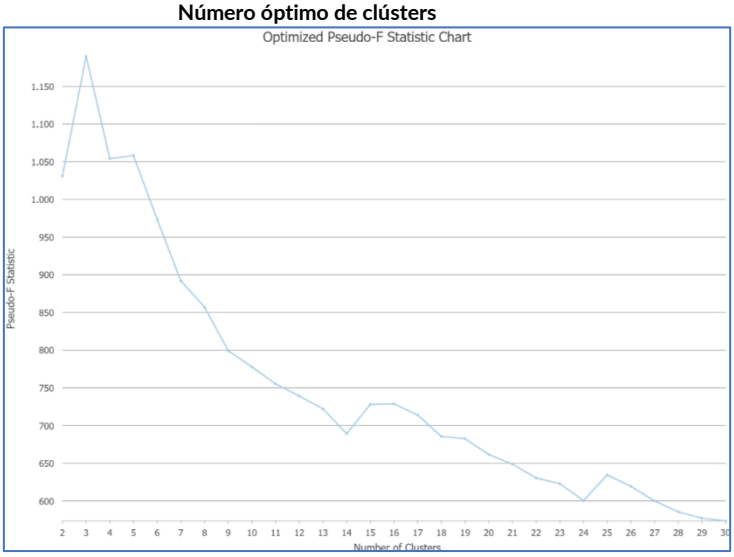
El primer análisis se ha realizado con las tramitaciones iniciadas antes del estado de alarma (desde el 1 de enero de 2019 al 18 de marzo de 2020).

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

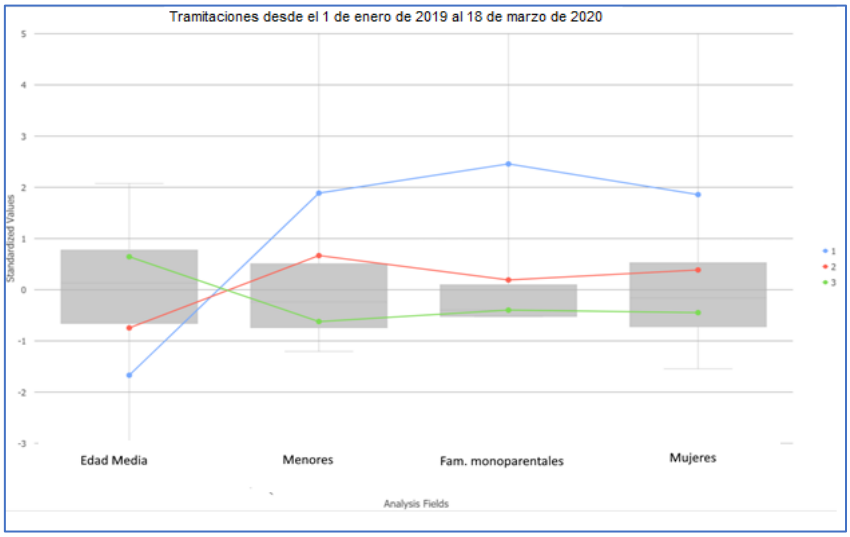
La elección del número de clústers a usar puede ser fijada de antemano, aunque también puede calcularse a través del análisis adecuado. La herramienta Multivariate Clustering permite lanzar varios procesos, iterando para cada número de clústers, y obtener el valor del indicador pseudo estadístico F -Pseudo-F- (puede verse más información sobre este indicador en este enlace). Este indicador mide para qué número de clústers se obtienen clústers más homogéneos internamente y, al mismo tiempo, más diferentes entre sí.

A partir de las variables sociodemográficas utilizadas en este análisis, el indicador Pseudo-F estima que el número óptimo de clústers es 3. Se presenta a continuación la salida del análisis:

Número óptimo de clústers



Clúster. Tramitaciones antes del estado de alarma



Clúster 1: mujeres jóvenes con menores a su cargo

Clúster 2: personas jóvenes, normalmente en pareja, con menores a su cargo

Clúster 3: personas mayores, sin menores a su cargo, de ambos sexos y que viven normalmente en pareja

De izquierda a derecha, la primera caja de la figura representa la edad media de las personas solicitantes, seguida del número de menores en el hogar, número de familias monoparentales y número de mujeres solicitantes. El boxplot del gráfico muestra la dispersión de los valores para cada variable en el dataset, con una escala estandarizada para todas ellas. Cada clúster, por tanto, expresa una configuración distinta de la población según las cuatro variables utilizadas.

En este caso particular, la configuración de los clústers puede interpretarse así:

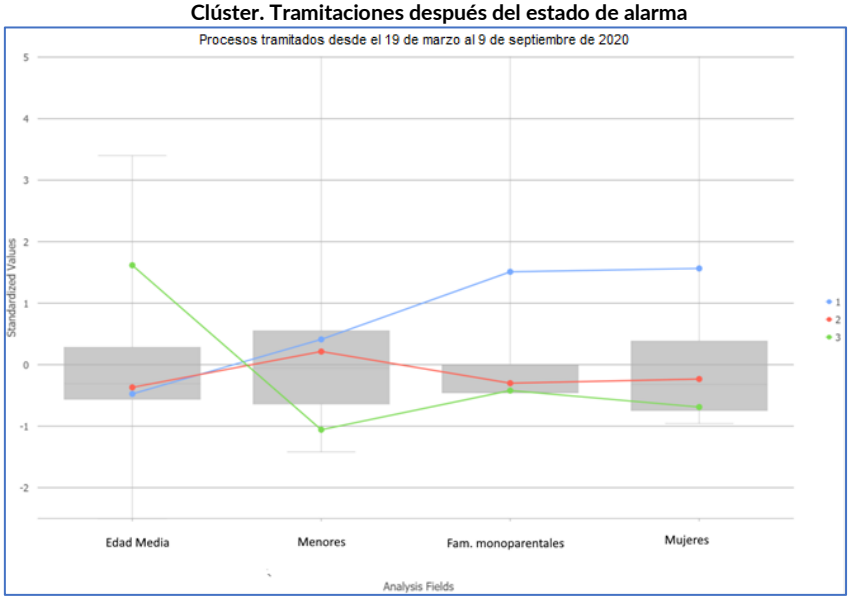
- El clúster 1 (en azul) representa a mujeres jóvenes con menores a su cargo.
- El clúster 2 (en rojo) representa a personas jóvenes, normalmente en pareja, con menores a su cargo.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

- El clúster 3 (en verde) representa a personas más mayores, sin menores a su cargo, de ambos sexos y que viven fundamentalmente en pareja.

El mismo análisis con las tramitaciones iniciadas después del estado de alarma ofrece una configuración de clústers muy similar, lo que indica que los grupos en los que se segmenta la población es consistente en todo el dataset y en distintos momentos del tiempo.

Clúster. Tramitaciones después del estado de alarma



Clúster 1: mujeres jóvenes con menores a su cargo

Clúster 2: personas jóvenes, normalmente en pareja, con menores a su cargo

Clúster 3: personas mayores, sin menores a su cargo, de ambos sexos y que viven normalmente en pareja

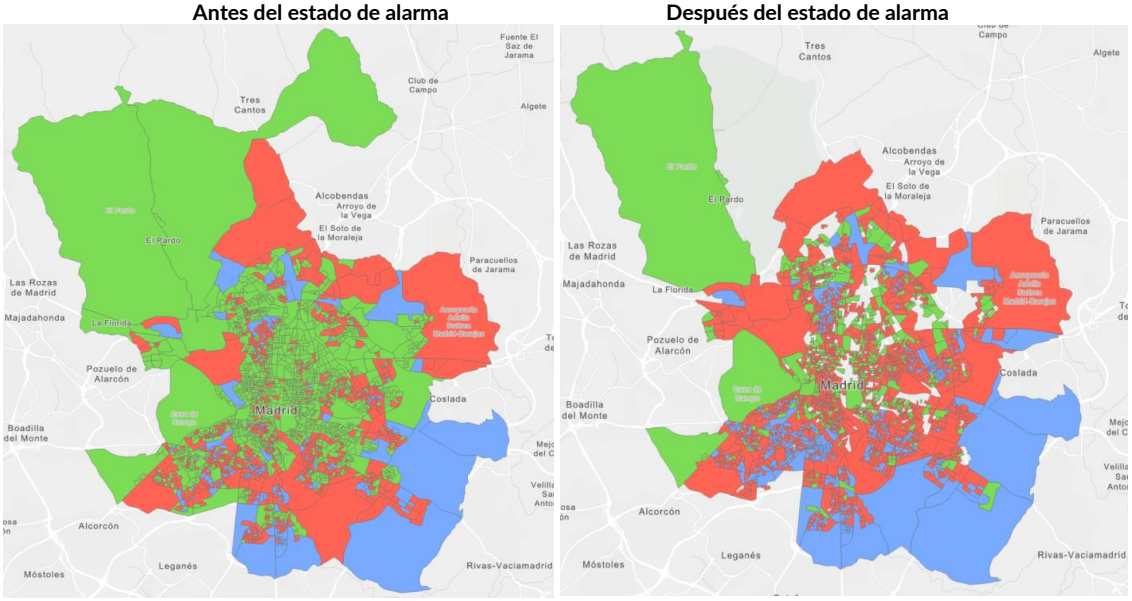
El principal cambio entre antes y después del estado de alarma es la proximidad de los valores de edad media y número de menores en los clústers 1 y 2.

La representación territorializada de las tramitaciones iniciadas antes (izquierda) y después (derecha) del estado de alarma muestran, en general, una reducción de la población mayor como demandante principal de ayudas (menor incidencia del clúster 3). Sería arriesgado concluir que la demanda de las personas de este último perfil ha disminuido; es posible que ,debido a la excepcionalidad de la situación, hayan decidido aplazar la solicitud.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Tramitaciones iniciadas registradas en CIVIS

Antes del estado de alarma **Después del estado de alarma**

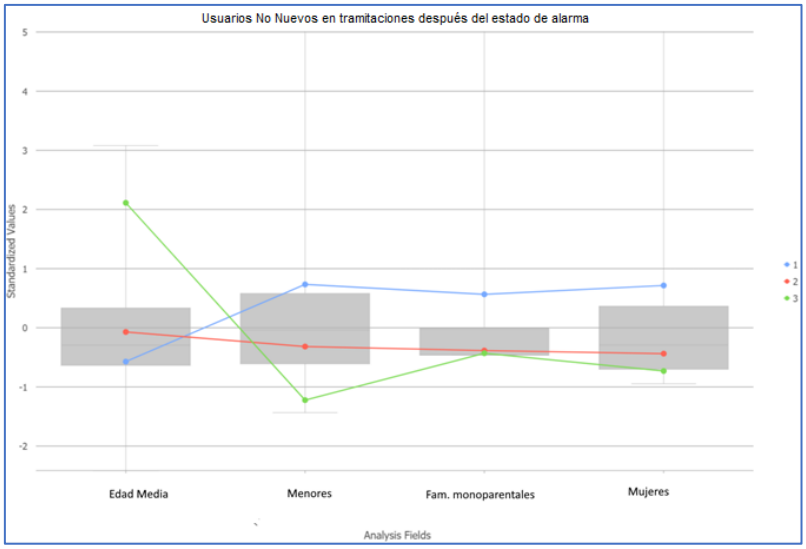


- Clúster 1: hogares de familia monoparentales con mujeres jóvenes con menores a su cargo.
- Clúster 2: hogares formados por personas jóvenes, normalmente en pareja, con menores a su cargo.
- Clúster 3: personas más mayores, sin menores a su cargo, de ambos sexos y que viven fundamentalmente en pareja.

El mismo proceso se realiza para las tramitaciones iniciadas después del estado de alarma (procesos tramitados desde el 19 de marzo al 9 de septiembre de 2020) pero distinguiendo entre personas usuarias nuevas -UN- y personas que tenían historia registrada en CIVIS con anterioridad al estado de alarma- UCH-.

Para personas usuarias con historial en CIVIS, el diagrama de cajas es muy similar al de antes del estado de alarma, sin que se hayan producido cambios significativos en los clústers.

Después del estado de alarma/Personas usuarias con historial -UCH-



Clúster 1: mujeres jóvenes con menores a su cargo

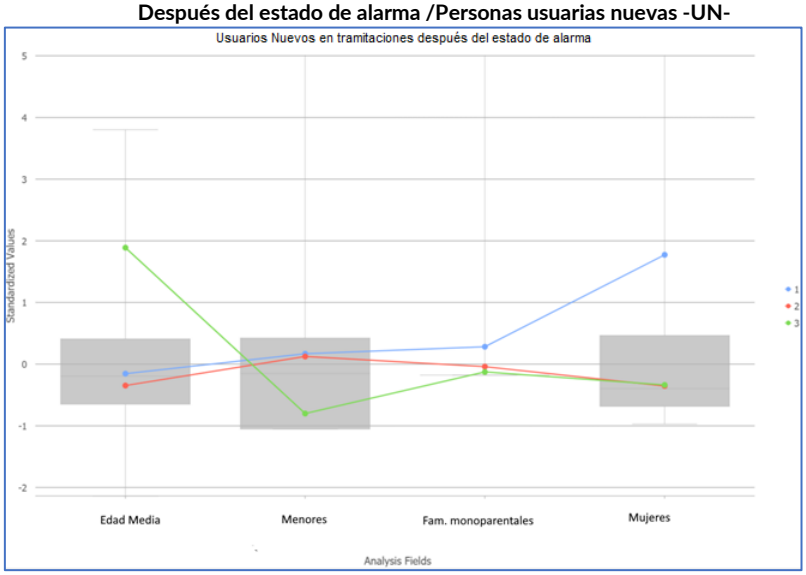
Clúster 2: personas jóvenes, normalmente en pareja, con menores a su cargo

Clúster 3: personas mayores, sin menores a su cargo, de ambos sexos y que viven normalmente en pareja

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

En el caso de personas usuarias nuevas se observa escasa variabilidad en cuanto al número de familias monoparentales en los tres clústers, y en cuanto a presencia de menores entre los clústers 1 y 2.

Después del estado de alarma /Personas usuarias nuevas -UN-



Clúster 1: mujeres jóvenes con menores a su cargo

Clúster 2: personas jóvenes, normalmente en pareja, con menores a su cargo

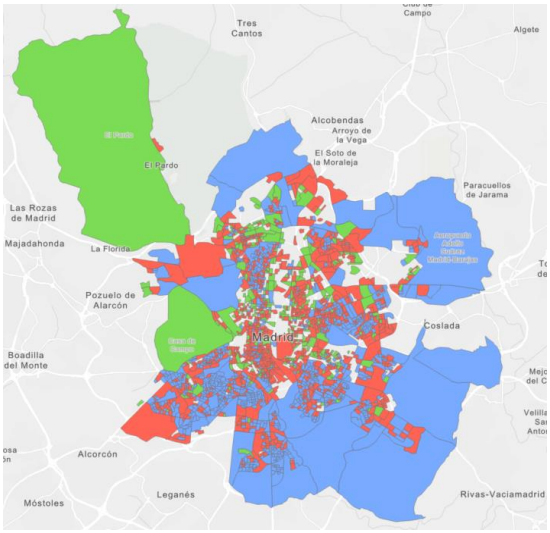
Clúster 3: personas mayores, sin menores a su cargo, de ambos sexos y que viven normalmente en pareja

La representación gráfica de las tramitaciones iniciadas después del estado de alarma de personas usuarias nuevas (derecha) y de personas usuarias con historial en CIVIS (izquierda) muestra un cambio muy significativo desde el clúster 1 (azul) hacia el clúster 2 (rojo). Es decir, donde tradicionalmente se observaban solicitudes de mujeres solteras, ahora se observan ,indistintamente, hombres o mujeres que pueden vivir en pareja.

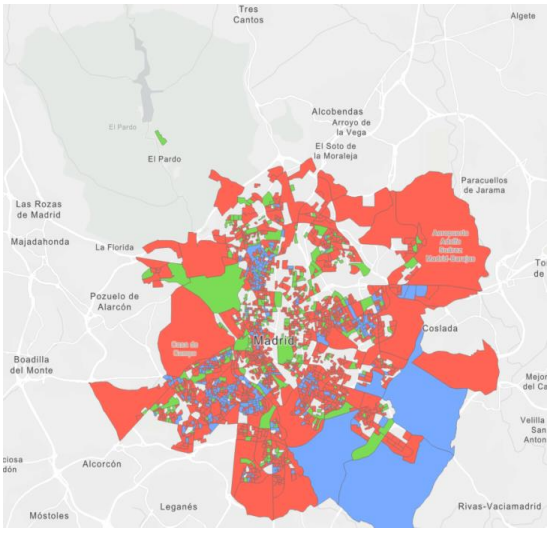
También se aprecian notables diferencias en el clúster 3 (personas mayores, sin menores a su cargo, de ambos sexos y que viven normalmente en pareja).

Tramitaciones iniciadas después del estado de alarma

Personas usuarias con historial -UCH-



Personas usuarias nuevas -UN-



- Clúster 1 hogares de familia monoparentales con mujeres jóvenes con menores a su cargo.
- Clúster 2 hogares formados por personas jóvenes, normalmente en pareja, con menores a su cargo.
- Clúster 3 personas más mayores, sin menores a su cargo, de ambos sexos y que viven fundamentalmente en pareja.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

2.3. Análisis de Hot Spots

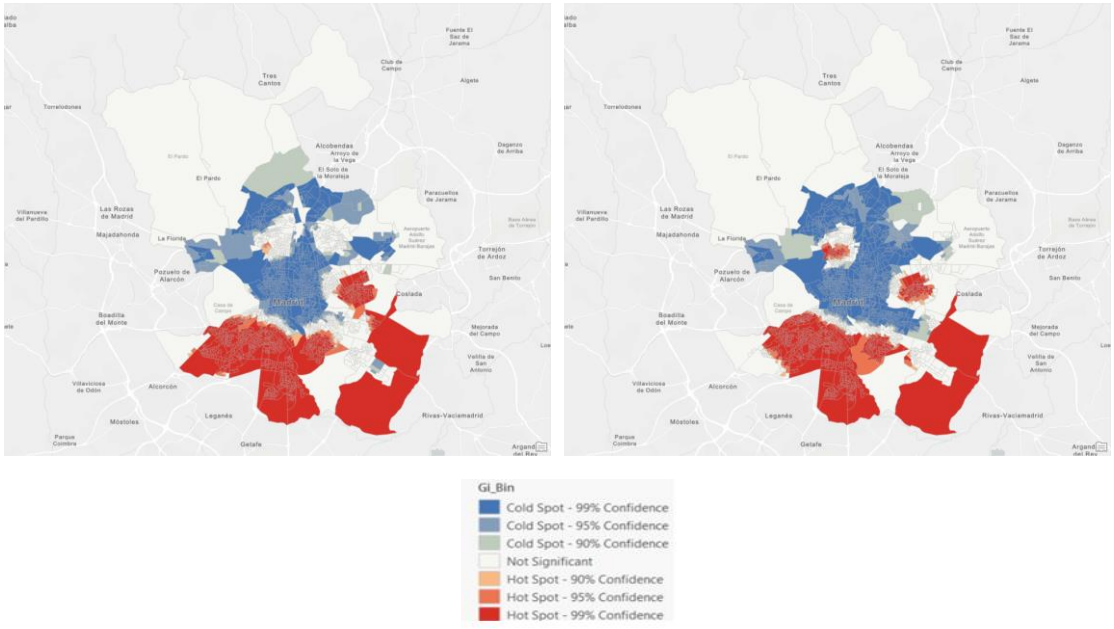
Aunque se disponga de mapas para explorar cómo varía la demanda en el territorio, es importante tener una visión estadística, y no sólo descriptiva, de la concentración de la misma.

La herramienta Optimized Hot Spot Analysis de ArcGIS permite identificar puntos calientes-hot spots- y puntos fríos estadísticamente significativos. Funciona examinando cada elemento dentro del contexto de los elementos vecinos. Un elemento con un valor alto puede ser de interés pero no ser un punto caliente estadísticamente significativo. Para ser un punto caliente estadísticamente significativo un elemento tendrá un valor alto y estará rodeado por otros elementos que también tendrán valores altos.

Para calcular los puntos calientes (hotspots) antes de la pandemia por covid19 se ha realizado este tipo de análisis a partir de las tramitaciones iniciadas antes del estado de alarma. Los puntos calientes se representan en una paleta de tonos rojos, los puntos fríos, en azul, y en blanco las zonas para las que no hay significación estadística.

El mismo análisis se ha realizado a partir de las tramitaciones iniciadas después del estado de alarma, de modo que se pueda establecer una comparativa entre ambos resultados. A continuación se ofrece la representación territorial del resultado de ambos análisis.

Hotspot. Antes del estado de alarma (izquierda) y después (derecha)

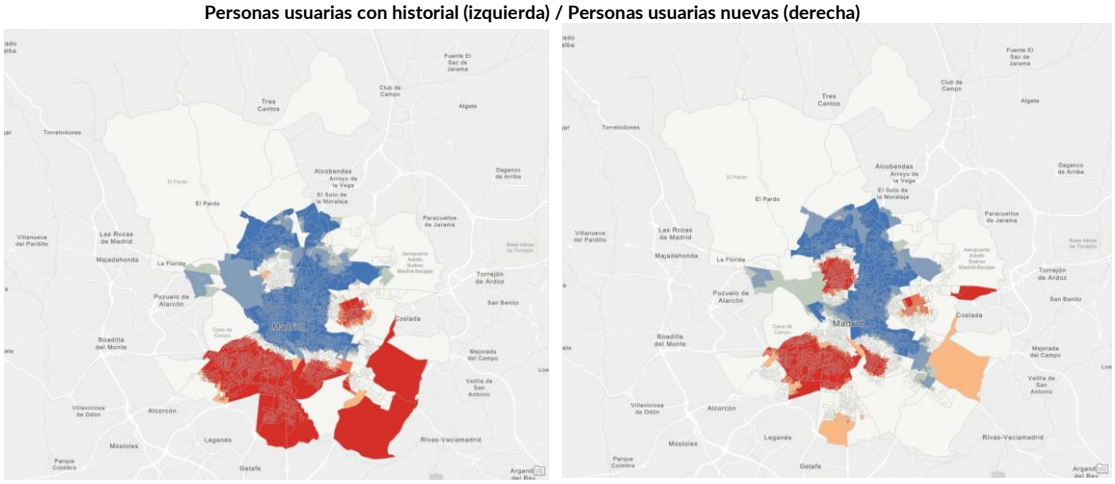


Se observa una distribución geográfica muy similar, con la aparición de una zona caliente en el área de Tetuán y una extensión del clúster de la zona sur (hacia Cuatro Vientos).

El mismo análisis se ha repetido con las tramitaciones iniciadas asociadas a las llamadas al 010, distinguiendo entre personas usuarias que ya tenían historia en CIVIS -UCH- y personas usuarias nuevas -UN-.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Hotspot. Después del estado de alarma.
Personas usuarias con historial (izquierda) / Personas usuarias nuevas (derecha)



De nuevo, se percibe la intensificación del foco de Tetuán y la diferente configuración de la región sur (se aportan algunas conclusiones adicionales sobre este estudio en el apartado 3).

El análisis de hotspot es sensible a la distancia escogida para el análisis, que determina el número de vecinos cercanos de cada entidad. Como no hay una distancia que, a priori, pueda ser considerada como apropiada para el análisis, se permitió que la herramienta calculara la adecuada. El método predilecto es la autocorrelación incremental, un proceso iterativo en el que se identifica, para distintos incrementos de distancia, la autocorrelación del dataset. La herramienta escoge la distancia a la que se ha calculado una mayor autocorrelación espacial, usando el método estadístico I de Moran. En estos todos los casos la banda es de 2.374,5147 metros (a esta distancia, los datos presentan una autocorrelación más fuerte). Se comprobó que la banda de distancia fuera la misma en todos los procesos, de modo que los resultados pudieran ser comparados entre sí.

2.4. Modelos explicativos

Partiendo de la capa de seccionado censal se ha generado un modelo que explique la demanda (llamadas al 010) en base a diferentes variables del territorio, tanto demográficas (población, edad, educación, estado civil, nacionalidad, hogares) como socioeconómicas (renta, paro, empleo, nivel económico).

Para estudiar la significancia de las variables y su colinealidad se utiliza la herramienta Exploratory regression, que evalúa todas las posibles combinaciones de una serie de variables y ofrece datos sobre la significancia de cada variable y de su relación con el resto, además de sugerir modelos de correlación lineal prometedores (es decir, qué combinaciones de las variables independientes ofrecen un mejor ajuste).

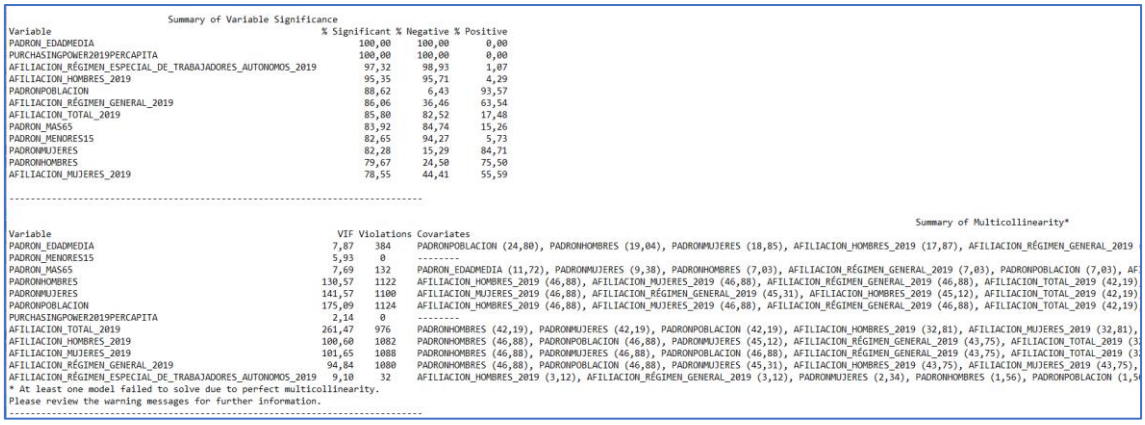
La limitación encontrada en este caso es la necesidad de comenzar con un conjunto de variables explicativas tan amplio como fuera posible. Aunque se pueda partir de la hipótesis de que la demanda proviene de un perfil demográfico concreto (clase baja, alto paro, familias con menores a cargo, con alto número de población nacida en el extranjero), ésta ha de ser contrastada a través de un análisis riguroso.

La Regresión Exploratoria no puede crear combinaciones de todas las posibles variables entre sí debido a limitaciones en la capacidad de cálculo. Por ello, se han ejecutado distintos procesos de Regresión Exploratoria con grupos de variables que se consideran de interés estadístico (variables de edad, variables de nacionalidad, variables de renta y paro, etcétera). El objetivo no es obtener modelos funcionales, sino examinar la significancia estadística a la hora de explicar la variable dependiente (el número de llamadas al 010) y descartar las variables que, por colinealidad, puedan ser redundantes y causar problemas a la hora de hacer estudios de regresión posterior.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Se ha comenzado planteando qué variables son las que se cree que puedan influir en la demanda. Estas son las variables que se incluyeron en la capa y cubren, conceptualmente:

- Edad de la población.
- Renta (ingresos/poder adquisitivo).
- Población extranjera.
- Paro y población laboral (afiliaciones).
- Nivel educativo.
- Composición del hogar y estado civil.



Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
PADRON_EDADMEDIA	100,00	100,00	0,00
PURCHASINGPOWER2019PERCAPITA	100,00	100,00	0,00
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2019	97,32	98,93	1,07
AFILIACION_HOMBRES_2019	95,35	95,71	4,29
PADRONPOBLACION	88,62	6,43	93,57
AFILIACION_RÉGIMEN_GENERAL_2019	86,06	36,46	63,54
AFILIACION_TOTAL_2019	85,80	82,52	17,48
PADRON_MAS65	83,92	84,74	15,26
PADRON_MENORES15	82,65	94,27	5,73
PADRONMUJERES	82,28	15,29	84,71
PADRONHOMBRES	79,67	24,50	75,50
AFILIACION_MUJERES_2019	78,55	44,41	55,59

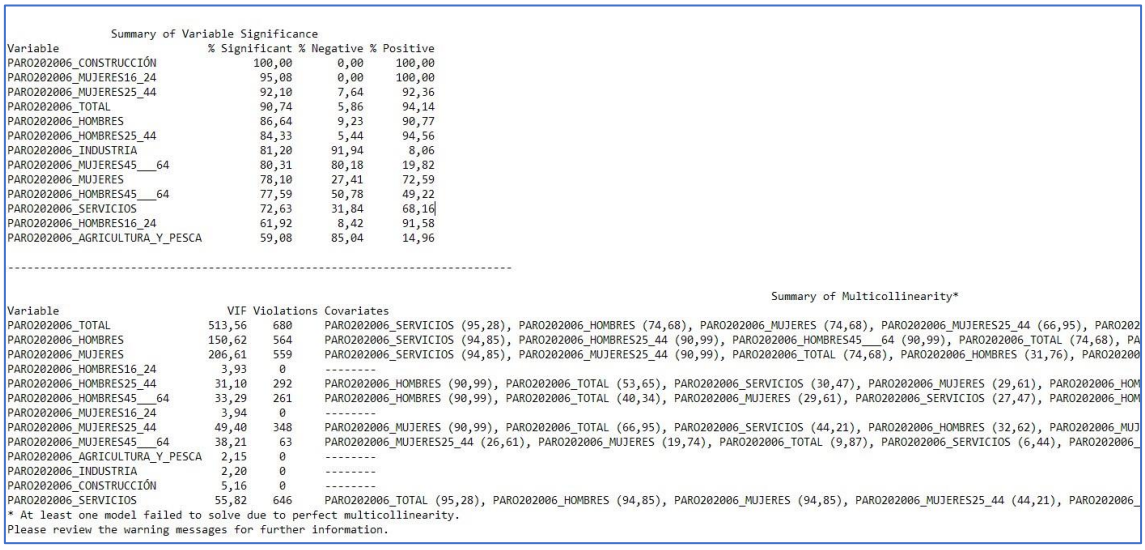
Summary of Multicollinearity*

Variable	VIF	Violations	Covariates
PADRON_EDADMEDIA	7,87	384	PADRONPOBLACION (24,80), PADRONHOMBRES (19,04), PADRONMUJERES (18,85), AFILIACION_HOMBRES_2019 (17,87), AFILIACION_RÉGIMEN_GENERAL_2019
PADRON_MENORES15	5,93	0	--------
PADRON_MAS65	7,69	132	PADRON_EDADMEDIA (11,72), PADRONMUJERES (9,38), PADRONHOMBRES (7,03), AFILIACION_RÉGIMEN_GENERAL_2019 (7,03), PADRONPOBLACION (7,03), AF
PADRONHOMBRES	130,57	1122	AFILIACION_HOMBRES_2019 (46,88), AFILIACION_MUJERES_2019 (46,88), AFILIACION_RÉGIMEN_GENERAL_2019 (46,88), AFILIACION_TOTAL_2019 (42,19)
PADRONMUJERES	141,57	1100	AFILIACION_MUJERES_2019 (46,88), AFILIACION_RÉGIMEN_GENERAL_2019 (45,31), AFILIACION_HOMBRES_2019 (45,12), AFILIACION_TOTAL_2019 (42,19)
PADRONPOBLACION	175,09	1124	AFILIACION_HOMBRES_2019 (46,88), AFILIACION_MUJERES_2019 (46,88), AFILIACION_RÉGIMEN_GENERAL_2019 (46,88), AFILIACION_TOTAL_2019 (42,19)
PURCHASINGPOWER2019PERCAPITA	2,14	0	--------
AFILIACION_TOTAL_2019	261,47	976	PADRONHOMBRES (42,19), PADRONMUJERES (42,19), PADRONPOBLACION (42,19), AFILIACION_HOMBRES_2019 (32,81), AFILIACION_MUJERES_2019 (32,81),
AFILIACION_HOMBRES_2019	100,60	1082	PADRONHOMBRES (46,88), PADRONPOBLACION (46,88), PADRONMUJERES (45,12), AFILIACION_RÉGIMEN_GENERAL_2019 (43,75), AFILIACION_TOTAL_2019 (3
AFILIACION_MUJERES_2019	101,65	1088	PADRONHOMBRES (46,88), PADRONMUJERES (46,88), PADRONPOBLACION (46,88), AFILIACION_RÉGIMEN_GENERAL_2019 (43,75), AFILIACION_TOTAL_2019 (3
AFILIACION_RÉGIMEN_GENERAL_2019	94,84	1080	PADRONHOMBRES (46,88), PADRONPOBLACION (46,88), PADRONMUJERES (45,31), AFILIACION_HOMBRES_2019 (43,75), AFILIACION_MUJERES_2019 (43,75),
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2019	9,10	32	AFILIACION_HOMBRES_2019 (3,12), AFILIACION_RÉGIMEN_GENERAL_2019 (3,12), PADRONMUJERES (2,34), PADRONHOMBRES (1,56), PADRONPOBLACION (1,5

* At least one model failed to solve due to perfect multicollinearity.
Please review the warning messages for further information.

En los reportes que genera la herramienta de Regresión Exploratoria se puede examinar la significancia de las variables a la hora de explicar la variable dependiente, y también dónde se detecta colinealidad. El parámetro VIF, cuando resulta ser mayor de 7.5, expresa una redundancia en las variables; dicho de otro modo, una variable con un VIF elevado está explicando lo que otras variables del modelo ya han explicado.

Se ha explorado fundamentalmente las variables del paro y afiliaciones. Se determinó que el paro en el sector Construcción es el más significativo, y su relación con la demanda es positiva en todos los casos. El paro en los sectores de Industria, Construcción y Agricultura y Pesca son los únicos que no presentan colinealidad con otras variables de paro, ni siquiera con las variables de paro total o por rangos de edad.



Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
PARO202006_CONSTRUCCIÓN	100,00	0,00	100,00
PARO202006_MUJERES16_24	95,08	0,00	100,00
PARO202006_MUJERES25_44	92,10	7,64	92,36
PARO202006_TOTAL	90,74	5,86	94,14
PARO202006_HOMBRES	86,64	9,23	90,77
PARO202006_HOMBRES25_44	84,33	5,44	94,56
PARO202006_INDUSTRIA	81,20	91,94	8,06
PARO202006_MUJERES45___64	80,31	80,18	19,82
PARO202006_MUJERES	78,10	27,41	72,59
PARO202006_HOMBRES45___64	77,59	50,78	49,22
PARO202006_SERVICIOS	72,63	31,84	68,16
PARO202006_HOMBRES16_24	61,92	8,42	91,58
PARO202006_AGRICULTURA_Y_PESCA	59,08	85,04	14,96

Summary of Multicollinearity*

Variable	VIF	Violations	Covariates
PARO202006_TOTAL	513,56	680	PARO202006_SERVICIOS (95,28), PARO202006_HOMBRES (74,68), PARO202006_MUJERES (74,68), PARO202006_MUJERES25_44 (66,95), PARO202
PARO202006_HOMBRES	150,62	564	PARO202006_SERVICIOS (94,85), PARO202006_HOMBRES25_44 (90,99), PARO202006_HOMBRES45___64 (90,99), PARO202006_TOTAL (74,68), PA
PARO202006_MUJERES	206,61	559	PARO202006_SERVICIOS (94,85), PARO202006_MUJERES25_44 (90,99), PARO202006_TOTAL (74,68), PARO202006_HOMBRES (31,76), PARO20200
PARO202006_HOMBRES16_24	3,93	0	--------
PARO202006_HOMBRES25_44	31,10	292	PARO202006_HOMBRES (90,99), PARO202006_TOTAL (53,65), PARO202006_SERVICIOS (30,47), PARO202006_MUJERES (29,61), PARO202006_HOM
PARO202006_HOMBRES45___64	33,29	261	PARO202006_HOMBRES (90,99), PARO202006_TOTAL (40,34), PARO202006_MUJERES (29,61), PARO202006_SERVICIOS (27,47), PARO202006_HOM
PARO202006_MUJERES16_24	3,94	0	--------
PARO202006_MUJERES25_44	49,40	348	PARO202006_MUJERES (90,99), PARO202006_TOTAL (66,95), PARO202006_SERVICIOS (44,21), PARO202006_HOMBRES (32,62), PARO202006_MUJ
PARO202006_MUJERES45___64	38,21	63	PARO202006_MUJERES25_44 (26,61), PARO202006_MUJERES (19,74), PARO202006_TOTAL (9,87), PARO202006_SERVICIOS (6,44), PARO202006_
PARO202006_AGRICULTURA_Y_PESCA	2,15	0	--------
PARO202006_INDUSTRIA	2,20	0	--------
PARO202006_CONSTRUCCIÓN	5,16	0	--------
PARO202006_SERVICIOS	55,82	646	PARO202006_TOTAL (95,28), PARO202006_HOMBRES (94,85), PARO202006_MUJERES (94,85), PARO202006_MUJERES25_44 (44,21), PARO202006_

* At least one model failed to solve due to perfect multicollinearity.
Please review the warning messages for further information.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Las afiliaciones de autónomos son las más significativas, seguidas por las del régimen general. Se observa multicolinealidad entre la mayoría de las variables de afiliación, excepto la correspondiente a autónomos. Mientras que la dependencia de la afiliación de autónomos es negativa (cuantos más autónomos, menos demanda), la de afiliación al régimen general es positiva (cuantos más afiliados, más demanda), porque la afiliación al régimen general depende directamente de la población y, a más población, más demanda.

Summary of Variable Significance

Variable	% Significant	% Negative	% Positive
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020	100,00	100,00	0,00
AFILIACION_RÉGIMEN_GENERAL_2020	90,00	0,00	100,00
AFILIACION_TOTAL_2020	88,89	44,44	55,56
AFILIACION_MUJERES_2020	70,00	20,00	80,00
AFILIACION_HOMBRES_2020	60,00	70,00	30,00

Summary of Multicollinearity*

Variable	VIF	Violations	Covariates
AFILIACION_TOTAL_2020	119,85	7	AFILIACION_HOMBRES_2020 (33,33), AFILIACION_MUJERES_2020 (33,33), AFILIACION_RÉGIMEN_GENERAL_2020
AFILIACION_HOMBRES_2020	59,54	8	AFILIACION_RÉGIMEN_GENERAL_2020 (44,44), AFILIACION_TOTAL_2020 (33,33), AFILIACION_MUJERES_2020 (3
AFILIACION_MUJERES_2020	58,63	8	AFILIACION_RÉGIMEN_GENERAL_2020 (44,44), AFILIACION_TOTAL_2020 (33,33), AFILIACION_HOMBRES_2020 (3
AFILIACION_RÉGIMEN_GENERAL_2020	47,94	8	AFILIACION_HOMBRES_2020 (44,44), AFILIACION_MUJERES_2020 (44,44), AFILIACION_TOTAL_2020 (33,33)
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020	4,07	0	--------

* At least one model failed to solve due to perfect multicollinearity.
Please review the warning messages for further information.

Se ha repetido el proceso de Regresión Exploratoria hasta haber utilizado todas las variables incluidas en el dataset. A partir de ahí, se ha limitado el conjunto inicial de variables a un subconjunto de las más significativas para las cuales no existe colinealidad.

- Padrón: menores de 15 años
- Padrón: mayores de 65
- Renta Per cápita
- Afiliación Régimen Especial de Trabajadores autónomos
- Paro: Agricultura y pesca
- Paro: Industria
- Paro: Construcción
- educación Secundaria
- educación universitaria
- educación. Analfabeto
- Sin educación
- Educación Primaria
- Solteros
- Casados
- Divorciados
- Tamaño de hogar (personas por hogar)
- Tipología H (Con niños clase social baja
- Tipología F (Con adolescentes y niños clase social baja)
- Tipología G (Con/Sin jóvenes clase social baja)
- Tipología P (pensionistas en pareja nivel bajo de estudios)
- Tipología Q (pensionistas solitarios nivel bajo de estudios)
- Tipología I (DINK Clase social baja)
- Pob. de África
- Pob. de América
- Pob. de Asia

Para completar el estudio de las variables significativas se ha usado un algoritmo de Machine Learning llamado Clasificación de Bosque Aleatorio (Random Forest). ArcGIS implementa este algoritmo en una herramienta llamada Random Forest Classification and Regression.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

El algoritmo crea árboles de decisión que permiten asignar un valor (o categoría) a un elemento en función de las variables que lo describen. La potencia del algoritmo estriba en que genera un conjunto de árboles (al que se llama bosque) para cada uno de los cuales se asigna un subconjunto aleatorio de datos de entrada y un subconjunto aleatorio de variables explicativas. Cuanto más alto sea el número de árboles, más preciso debería ser el modelo. Como cada árbol recibe sólo un subconjunto aleatorio de los datos, los errores y los sesgos tienden a corregirse para el conjunto del bosque. Además, se reserva un 10% de los datos de entrada para calibrar la precisión del modelo.

Lo que se obtiene es un resultado aparentemente robusto:

---------- Model Out of Bag Errors ----------		
Number of Trees	250	500
MSE	87,889	87,440
% of variation explained	83,488	83,573

---------- Top Variable Importance ----------		
Variable	Importance	%
Pob. de America	322471,11	28
Paro202006_Construcción	197610,86	17
Paro202006_Servicios	131494,95	11
Pob. de Africa	85169,67	7
Renta Per Capita año 2019	74835,57	6
Paro202006_Mujeres16_24	55307,20	5
Pob. de Asia	40821,66	4
Paro202006_Hombres16_24	33599,01	3
Padron. Menores de 15 años	32932,67	3
EducacionPrimaria	21145,38	2
SinEducacion	20220,50	2
EducacionUniversitaria	15789,36	1
Paro202006_Industria	13945,80	1
Afiliacion_Régimen_Especial_de_Trabajadores_Autonomos_2020	13115,65	1
EducacionAnalfabeto	13032,59	1
Estado Civil Soltero	12922,82	1
Paro202006_Agricultura_y_pesca	11569,09	1
Tamaño del Hogar	8702,22	1
Estado Civil Casado	7275,27	1
Padron. Mayores de 65 años	7108,47	1

Un proceso de machine learning es tan bueno como la calidad de las muestras usadas para entrenarlo. Dado que solo se dispone de un conjunto de entidades con un solo conjunto de valores, no se puede garantizar que el modelo pueda predecir con precisión. Sin embargo, la principal utilidad de este algoritmo en este caso no es generar un modelo predictivo final, sino obtener un nuevo conjunto de variables relevantes para explicar la demanda de una forma independiente a como se obtuvieron antes. Si dos métodos distintos encuentran resultados generales, se puede concluir con mayor certeza qué variables serán más significativas a la hora de explicar la demanda.

De este modo, podemos combinar los resultados de la regresión exploratoria y del modelo de Random Forest para obtener un conjunto de variables que tengan significancia a la hora de explicar la variable dependiente.

Summary of Variable Significance			
Variable	% Significant	% Negative	% Positive
POBAMERICA	100,00	0,00	100,00
POBASIA	100,00	0,00	100,00
PARO202006_MUJERES16_24	100,00	0,00	100,00
PARO202006_CONSTRUCCIÓN	100,00	0,00	100,00
PARO202006_SERVICIOS	99,99	0,00	100,00
POBAFRICA	99,75	0,00	100,00
EDUCACIONPRIMARIA	98,55	0,00	100,00
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020	97,30	97,29	2,71
PARO202006_HOMBRES16_24	92,08	0,65	99,35
RENTA2019PERCAPITA	90,80	99,47	0,53
ESTADOCIVILSOLTERO	81,79	56,61	43,39
ESTADOCIVILCASADO	81,67	88,68	11,32
TAMAÑODEHOGAR	64,95	52,95	47,05
SINEDUCACION	59,73	8,01	91,99
PADRON_MENORES15	51,63	50,78	49,22
ESTADOCIVILDIVORCIADO	50,23	59,14	40,86

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

El proceso no detectó colinealidad entre ninguna de las variables, y el índice VIF es suficientemente bajo en todas ellas (no indica redundancia entre variables).

El siguiente paso es estudiar cómo funcionan estas variables en un modelo de regresión lineal. Para ello se ha utilizado la herramienta *Generalized Linear Regression* (GLR).

El modelo de regresión generalizado es un ajuste de mínimos cuadrados que ofrece opciones para trabajar con datos que no se distribuyen de forma normal. En este caso se ha trabajado con recuentos (número de personas solicitantes), por lo que se ha usado la distribución de Poisson en la herramienta Generalized Weighted Regression. El resultado del modelo ha sido el siguiente:

Variable	Coefficient [a]	StdError	z-Statistic	Probability [b]	VIF [c]
Intercept	4,467547	0,060077	74,363160	0,000000*	--------
PADRON_MENORES15	-0,000825	0,000073	-11,363280	0,000000*	3,651837
RENTA2019PERCAPITA	-0,000082	0,000002	-43,871555	0,000000*	4,551341
SINEDUCACION	-0,000760	0,000082	-9,244243	0,000000*	2,302813
EDUCACIONPRIMARIA	-0,000270	0,000064	-4,233667	0,000023*	2,099671
ESTADOCIVILSOLTERO	0,000535	0,000040	13,245602	0,000000*	5,805492
ESTADOCIVILCASADO	-0,000081	0,000035	-2,333915	0,019600*	3,221353
ESTADOCIVILDIVORCIADO	0,000073	0,000126	0,579975	0,561931	1,338556
TAMAÑODEHOGAR	-0,284963	0,015913	-17,907723	0,000000*	1,487389
POBAFRICA	0,000076	0,000147	0,514407	0,606968	2,205347
POBAMERICA	0,003626	0,000073	49,819599	0,000000*	2,286663
POBASIA	0,001598	0,000086	18,625513	0,000000*	1,555103
AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020	-0,003884	0,000288	-13,476755	0,000000*	3,932788
PARO202006_HOMBRES16_24	0,006332	0,001720	3,682687	0,000231*	3,184893
PARO202006_MUJERES16_24	0,005647	0,001416	3,988294	0,000067*	3,099946
PARO202006_CONSTRUCCIÓN	-0,001075	0,001204	-0,893272	0,371712	3,909878
PARO202006_SERVICIOS	0,007960	0,000266	29,905262	0,000000*	6,683089

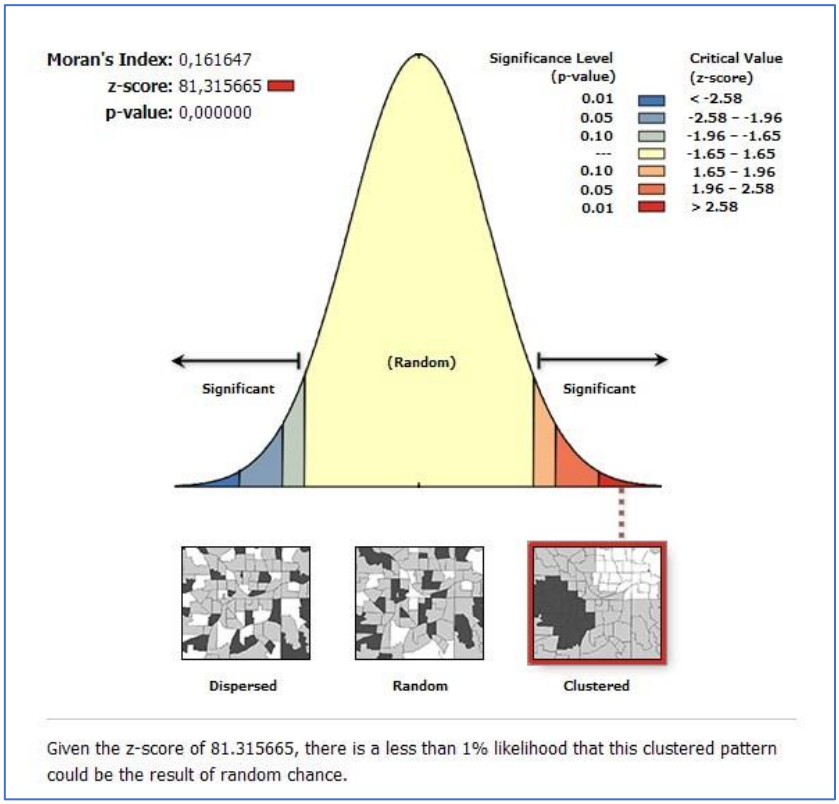
GLR Diagnostics

Input Features:	SeccionadoCensal	Dependent Variable:	PERSONAS_QUE_HAN_LLAMADO
Number of Observations:	2443	Akaike's Information Criterion (AICc) [d]:	19008,000000
Average Count:	22,438395	Deviance Explained [e]:	0,828091
Joint Wald Statistic [f]:	39975,924427	Prob(>chi-squared), (16) degrees of freedom:	0,000000*

Se consigue explicar casi un 83% de la varianza de la variable dependiente. En general, los factores de probabilidad de todas las variables son bajos, excepto para el Estado Civil Divorciado, la Población de África y el Paro en el sector Construcción. Una probabilidad mucho mayor que cero indica que no se puede descartar que la variable en cuestión no sea irrelevante para explicar la variable dependiente. La estadística Conjunta de Wald también tiene un valor muy alto como para garantizar la significancia del modelo; esto suele indicar que faltan factores clave para explicar la variable dependiente, y en particular factores territoriales que hagan que el modelo pueda ajustarse a toda el área de estudio de forma homogénea.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Para investigar sobre ello, se ha estudiado los residuos. Si los residuos no se distribuyen aleatoriamente en el territorio, el modelo no funciona igual para cada zona. Esto suele indicar que existe un factor geográfico que impacta en la capacidad de explicar la variable dependiente. Puede comprobarse numéricamente mediante el test de autocorrelación espacial global I de Moran (implementado en ArcGIS en la herramienta Spatial Autocorrelation (Global Moran's I)). Este test comprueba la autocorrelación espacial de los resíduos del modelo de GLR indicando si existen clústers estadísticamente significativos o si el conjunto de datos está disperso.



Esto significa que el modelo calculado por el proceso de GLR presenta sesgos locales.

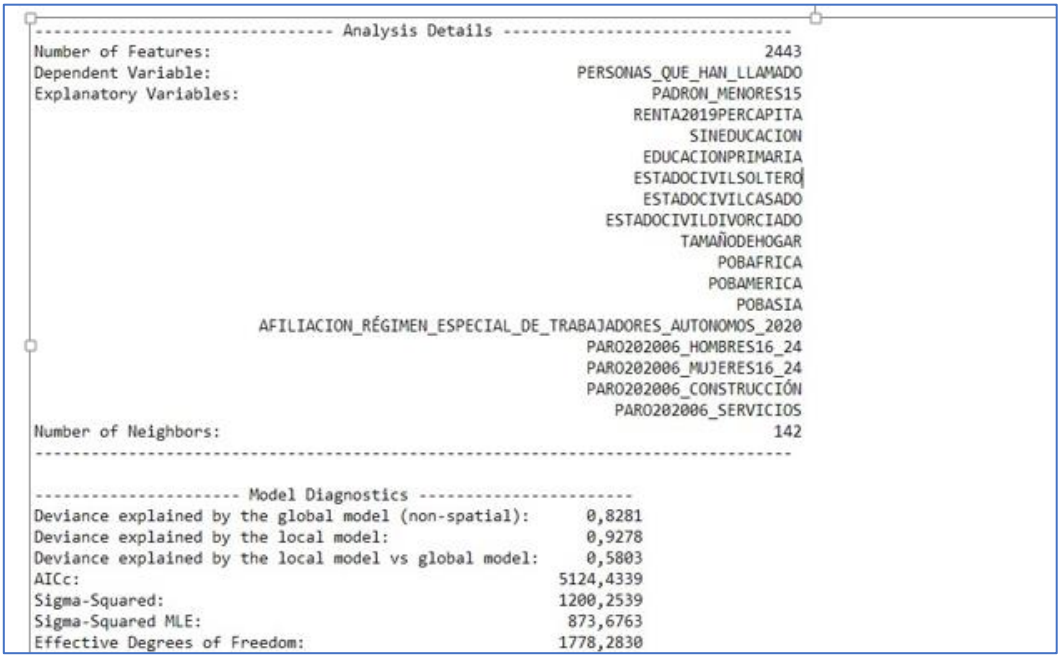
Para mejorar el modelo, se ha optado por usar un método de regresión geográfica o *GWR* (los parámetros de este modelo se pueden consultar en la ayuda de la herramienta y en este link se puede ver cómo funciona).

El modelo *GWR* (Geographic Weighted Regression) ajusta un modelo global similar generado con la herramienta GLR. Posteriormente, ajusta los coeficientes del modelo para cada entidad de entrada (sección censal) en función de sus vecinos cercanos. La herramienta calcula el número ideal de vecinos cercanos mediante un proceso iterativo que comprueba la correlación a cada banda de distancia.

De este modo, GWR permite ajustar localmente el modelo teniendo en cuenta el entorno geográfico de cada entidad y logrando un mejor ajuste en todo el territorio.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

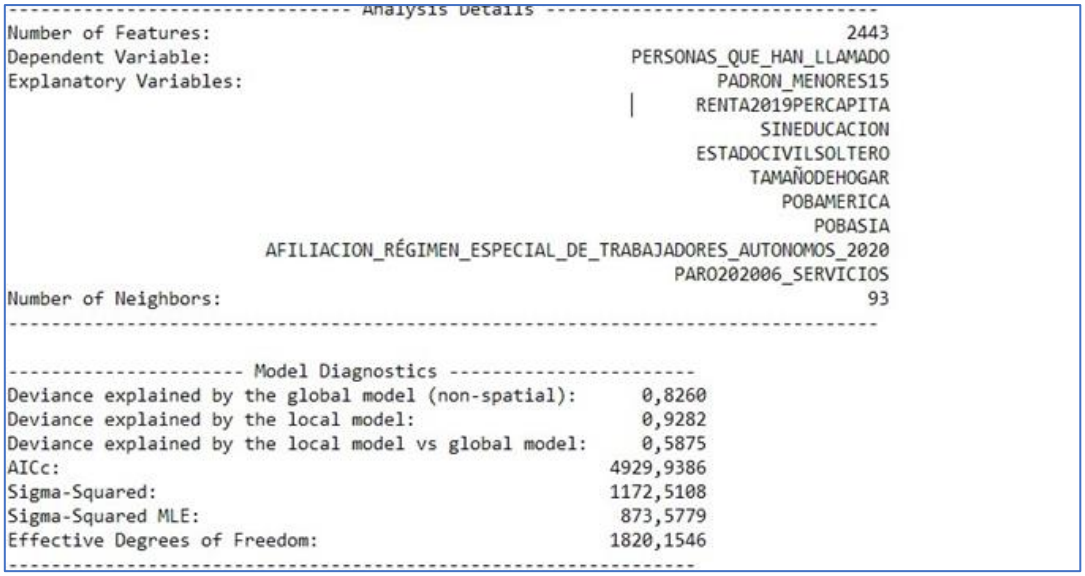
Se ha ejecutado el modelo usando las variables y obteniendo el modelo de diagnostico siguientes:



```
------------------------------ Analysis Details ------------------------------
Number of Features:                                                      2443
Dependent Variable:                                  PERSONAS_QUE_HAN_LLAMADO
Explanatory Variables:                                       PADRON_MENORES15
                                                        RENTA2019PERCAPITA
                                                              SINEDUCACION
                                                          EDUCACIONPRIMARIA
                                                        ESTADOCIVILSOLTERO
                                                         ESTADOCIVILCASADO
                                                      ESTADOCIVILDIVORCIADO
                                                            TAMAÑODEHOGAR
                                                                POBAFRICA
                                                               POBAMERICA
                                                                 POBASIA
                                 AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020
                                                       PARO202006_HOMBRES16_24
                                                       PARO202006_MUJERES16_24
                                                       PARO202006_CONSTRUCCIÓN
                                                          PARO202006_SERVICIOS
Number of Neighbors:                                                      142
------------------------------------------------------------------------------

---------------------- Model Diagnostics -----------------------
Deviance explained by the global model (non-spatial):      0,8281
Deviance explained by the local model:                     0,9278
Deviance explained by the local model vs global model:     0,5803
AICc:                                                   5124,4339
Sigma-Squared:                                          1200,2539
Sigma-Squared MLE:                                       873,6763
Effective Degrees of Freedom:                           1778,2830
```

En este caso, se logra explicar casi el 93% de la varianza de la variable dependiente. Se comprobó si el modelo mejoraba suprimiendo las variables que parecían más dudosas durante el GLR: Estado Civil Divorciado, Población de África y Paro en el sector Construcción, así como las variables de paro por edad.



```
------------------------------ Analysis Details ------------------------------
Number of Features:                                                      2443
Dependent Variable:                                  PERSONAS_QUE_HAN_LLAMADO
Explanatory Variables:                                       PADRON_MENORES15
                                                        RENTA2019PERCAPITA
                                                              SINEDUCACION
                                                        ESTADOCIVILSOLTERO
                                                            TAMAÑODEHOGAR
                                                               POBAMERICA
                                                                 POBASIA
                                 AFILIACION_RÉGIMEN_ESPECIAL_DE_TRABAJADORES_AUTONOMOS_2020
                                                          PARO202006_SERVICIOS
Number of Neighbors:                                                       93
------------------------------------------------------------------------------

---------------------- Model Diagnostics -----------------------
Deviance explained by the global model (non-spatial):      0,8260
Deviance explained by the local model:                     0,9282
Deviance explained by the local model vs global model:     0,5875
AICc:                                                   4929,9386
Sigma-Squared:                                          1172,5108
Sigma-Squared MLE:                                       873,5779
Effective Degrees of Freedom:                           1820,1546
------------------------------------------------------------------
```

La varianza explicada crece muy ligeramente, y también desciende ligeramente el índice AICc, que es un medidor relativo de la significancia del modelo. Por tanto, esta última versión del modelo se ha elegido como resultado final.

Como conclusión, se considera que es necesario disponer de más datos para poder modelizar adecuadamente la demanda. En particular, se considera que el factor más crítico (el paro) no está adecuadamente expresado en los datos disponibles.Los últimos datos de paro no describen con precisión lo ocurrido durante el estado de alarma, en particular lo que pudo suceder en sectores críticos como la hostelería o los servicios, ni el impacto de los ERTEs en la economía de las familias. Por tanto, este modelo está tratando de describir un suceso muy insólito y acotado en el tiempo con datos estáticos que reflejan una tendencia más general de la población.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

Aunque no se puede garantizar que el modelo calculado sea extrapolable a otros periodos de tiempo, los datos usados en el proyecto son, a nivel sociodemográfico y socioeconómico, los mejores disponibles. Por tanto, se considera que el modelo describe la demanda de la mejor manera posible en base a los atributos del territorio que se han elegido como relevantes.

Todas las herramientas usadas, y en particular la Geographic Weighted Regression, permiten incluir una capa de variables nuevas que devuelvan una capa de predicción. De este modo, pueden usarse las herramientas para generar un modelo explicativo y producir una capa predictiva a partir de un conjunto de variables distinto.

3. Conclusiones generales del estudio

De los resultados obtenidos a través de los procesos de análisis se pueden extraer las siguientes conclusiones:

3.1. Distribución geográfica de la población vulnerable

La distribución geográfica de la población vulnerable no ha cambiado significativamente.

Si bien la pandemia por Covid-19 ha incrementado la demanda en zonas concretas de la ciudad, la distribución geográfica de la población demandante es similar a la que existía antes de la pandemia.

Los mapas de hotspots muestran claramente estos patrones de distribución, y presentan grandes similitudes cuando se comparan los datos antes y después del estado de alarma y los de los perfiles (personas usuarias nuevas y las que ya tenían historial en CIVIS) de las tramitaciones iniciadas después del estado de alarma.

Es importante remarcar, no obstante:

- Que los perfiles de la nueva persona solicitante tienden a estar más concentrados, lo cual significa que las áreas más vulnerables son más pequeñas e intensas.
- Que, como efecto de la pandemia, se ha intensificado la demanda en zonas que, previamente, no eran tan significativas: en particular en la zona de Berruguete-Valdezarza.

La limitación del estudio es que los resultados del análisis de hotspots pueden variar significativamente según la distancia a la que se comparen vecinos cercanos. Por esta razón se ha permitido que el proceso estime automáticamente la banda de distancia a la cual se obtiene una mayor autocorrelación. Como todos los procesos de análisis de hotspots se han calculado con la misma banda, son consistentes entre sí, y se considera que las conclusiones son bastante fiables.

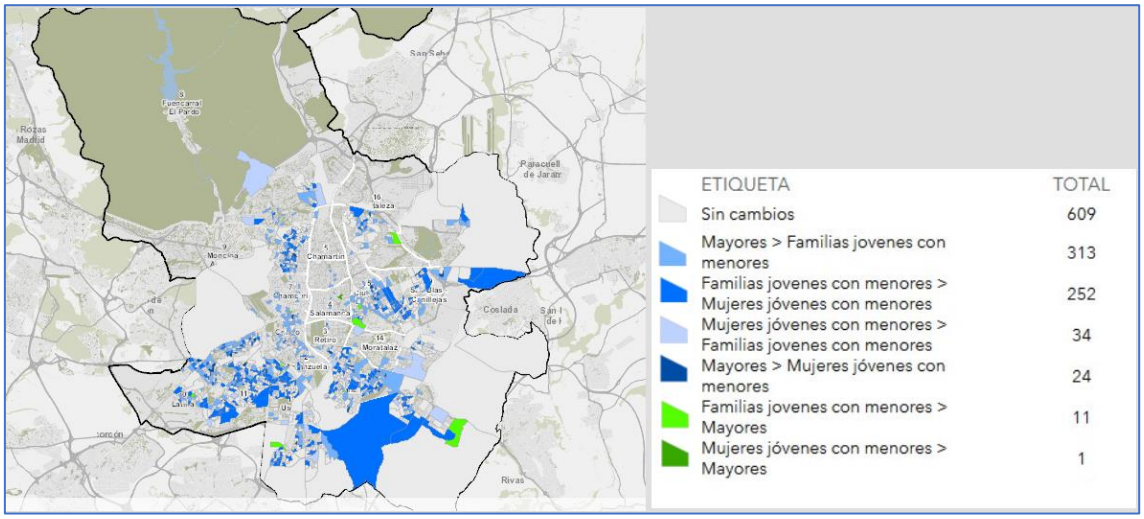
Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

3.2. Perfil demográfico del demandante

El perfil del demandante parece haber cambiado.

Los análisis sugieren que la nueva demanda surgida a raíz de la pandemia por covid 19 ha evolucionado hacia una población en la que predominan las familias monoparentales (fundamentalmente mujeres solteras).

El análisis de clústers revela que ha habido cambios en 645 de las secciones censales analizadas de las 1244 que tienen, al menos, 10 tramitaciones antes y después del estado de emergencia (55%). El cambio significa que el perfil demográfico ha cambiado lo suficiente como para que el sistema lo asocie a un nuevo clúster.



De las secciones censales en las que ha habido cambios, más de un 49% corresponden a zonas en las que el perfil ha cambiado de personas mayores a familias jóvenes con menores y el 39% se corresponden a zonas en las que el perfil ha cambiado de familias jóvenes con menores a familias monoparentales con menores a su cargo.

La principal objeción a este análisis es que en algunas secciones censales el número de registros puede ser pequeño. No obstante, las zonas con poca demanda son zonas, por ello, menos interesantes para el estudio y se han eliminado del mismo para evitar añadir ruido. Además, la consistencia en la generación de clústers independientemente de la muestra utilizada (dataset completo, antes y después de la pandemia, personas usuarias nuevas o no) sostiene la validez del proceso.

3.3. Modelización y predicción de la demanda

El análisis de modelización realizado para explicar la demanda a partir de parámetros territoriales revela una fuerte dependencia datos como el paro, el poder adquisitivo y el nivel educativo.

Cabe la pena pensar si es posible explicar de forma confiable un fenómeno tan extremo y acotado en el tiempo como la pandemia causada por el SARS-CoV-19 a partir de datos demográficos mucho más estáticos. Pero los análisis previos muestran que la demanda, a nivel territorial, se distribuye esencialmente igual que antes de la pandemia. Por tanto, no hay razón para rechazar la validez del modelo explicativo ni su capacidad para expresar la demanda a lo largo del territorio.

Se ha usado un modelo ponderado geográficamente, ya que no fue posible ajustar un modelo global que se comportara de forma homogénea en toda la ciudad. Al ajustar los coeficientes localmente a partir de los vecinos cercanos, se consigue explicar un 90% de la varianza de la variable dependiente.

No se han realizado predicciones a 3, 6 y 12 meses por falta de datos históricos estables que permitan obtener conclusiones fiables. Para poder realizar un análisis predictivo se estiman varias estrategias posibles:

- La extrapolación de resultados históricos: En este caso particular, los datos de demanda disponibles pertenecen a un momento temporal muy concreto, de algo menos de tres meses, que corresponden

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

a una situación muy excepcional. ArcGIS dispone de herramientas de análisis espaciotemporal para realizar predicciones a partir de series históricas, pero se considera que este análisis sólo podrá hacerse más adelante, cuando existan datos suficientes y puedan identificarse patrones de cambio estacional significativos.

- La comparación con eventos similares: La primera aproximación al problema fue comparar la presente emergencia con la crisis económica acaecida entre 2008 y 2012, aproximadamente. Aunque se disponía de algunos datos demográficos de tales fechas, no se disponía de información de demanda para poder realizar modelos de correlación. Además, parece dudoso que la crisis de entonces sea comparable a la presente, mucho más repentina y por causas enteramente distintas.

Si se asume que el modelo explicativo obtenido en el presente estudio es válido y seguirá siéndolo en los próximos meses, sería posible obtener predicciones si se dispusiera de proyecciones de datos demográficos. Como se ha comentado, el paro es, quizá, la variable más cambiante de las que tienen alta significancia en el modelo.

Sin embargo, realizar modelos de evolución del paro para 3, 6 y 12 meses excede del alcance de este proyecto. Si el Instituto Nacional de Estadística o el Área de Estadística del Ayuntamiento de Madrid proporcionaran dichas proyecciones, las mismas herramientas de ArcGIS usadas en este estudio para obtener modelos explicativos podrían ser usadas para calcular capas predictivas a partir de estos valores proyectados.

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es

MADRID

esri España
THE SCIENCE OF WHERE™

Paseo de la Chopera, 41 – 1ª pta. 28045 Madrid
dginnovacion@madrid.es