



Abrir residencias en España no es solo una decisión empresarial: es una decisión social

Equipo Sandía

Este informe fue compilado por última vez a 5 de febrero de 2026.

Abrir residencias en España no es solo una decisión empresarial: es una decisión social. El envejecimiento demográfico proyecta que más del 30 % de la población española supere los 65 años para 2050, mientras que el déficit actual supera las 13,800 plazas solo en Galicia. ¿Dónde ubicar 1,000 nuevas residencias sin fallar en el intento? El algoritmo SEED (*Socio-Economic and Environmental Distribution*) responde a este desafío procesando 36,000+ secciones censales del INE mediante un modelo multicriterio que integra: demanda residencial (45 %), viabilidad económica (40 %) y saturación territorial (15 %). Validado con correlación $r=0.882$ respecto a la distribución real del sector y verificado con ocupaciones del 75.4 % en ubicaciones top, SEED transforma datos en decisiones operativas para un reto socioeconómico crítico.

Desafío demográfico y social

El problema real. Según el Instituto Nacional de Estadística (INE), España enfrenta uno de los mayores desafíos socioeconómicos del siglo XXI: se estima que para 2050 más del 30 % de la población superará los 65 años. Este fenómeno genera una creciente demanda de infraestructuras para el cuidado de personas mayores. La decisión de dónde ubicar las residencias no es trivial. Factores como la densidad poblacional, el poder adquisitivo y la accesibilidad geográfica deben considerarse simultáneamente.

Público objetivo. La respuesta no es obvia: para diseñar un algoritmo de localización efectivo, primero debemos entender quién toma realmente la decisión de ingreso. El público objetivo de una residencia no es solo el residente anciano, sino el sistema familiar completo. No vendemos plazas a ancianos, vendemos alivio y tranquilidad a familias saturadas.

Nuestra solución. Para ello, mediante la previa búsqueda de conocimiento y de información, y el desarrollo de un *dataframe* completo a través de diversos usos de fuentes de información, es posible identificar qué tres pilares sustentan la decisión de localización, con pesos diferenciados según su relevancia estratégica: factores de demanda, capacidad económica, y entorno.

Metodología. Arquitectura de SEED

El algoritmo SEED se estructura en cuatro capas jerárquicas que procesan y combinan diferentes dimensiones del problema.

Capa 1: Base Territorial. Define el espacio de decisión completo utilizando las 36,000+ secciones censales del INE como unidades geográficas básicas. Cada sección incluye: identificador único (`id_seccion`), coordenadas geográficas (`latitud`, `longitud`), y código provincial para agregaciones administrativas.

Capa 2: Demanda Residencial (Peso: 45 %). Esta capa evalúa el potencial de demanda local mediante tres indicadores complementarios:

- **Figure of Merit (F-of-M) - Peso interno: 65 %:** Métrica propietaria que evalúa la idoneidad demográfica de cada sección con respecto a nuestra pirámide poblacional ideal. Valores más bajos indican mayor idoneidad. Se normaliza y se invierte para que *scores* altos = mejor ubicación.
- **Índice de Dependencia - Peso interno: 10 %:** Mide la proporción de población mayor con grado de dependencia III, pues estos son, aunque no únicos, parte de la población afectada. Valores más altos indican mayor demanda potencial. La ponderación interna es baja debido a que a pesar de ser una variable relevante, puede generar ruido dentro del algoritmo debido a que los datos encontrados son autonómicos y no ligados al censo particular. Se normalizan los valores a $[0,1]$ para amplificar diferencias sutiles.

- **Densidad Poblacional (hab/km²) - Peso interno: 25 %:** El factor clave para el modelo es que un tercio de la población urbana será mayor de 60 años a largo plazo, entrando en nuestros rangos de interés. La ubicación en zonas urbanas beneficia la logística de la residencia, permitiendo así mayor accesibilidad y mayor concentración de recursos. La normalización es crítica debido al rango extremadamente amplio.

Capa 3: Viabilidad Económica (Peso: 40 %). Esta capa utiliza la renta media por hogar como *proxy* de la capacidad de pago, para así cumplir requisito de no morosidad. A diferencia de las otras variables NO se normaliza linealmente, sino que se aplica una función asimétrica:

$$\text{Score}(\text{Renta}) = \begin{cases} 0, & \text{Renta} \leq 40,000 \\ \left(\frac{\text{Renta} - 40,000}{32,000} \right)^{0,7}, & 40,000 < \text{Renta} \leq 72,000 \\ \frac{1}{1 + 0,5 \left(\frac{\text{Renta} - 72,000}{32,000} \right)^{1,5}}, & \text{Renta} > 72,000 \end{cases}$$

*El precio medio mensual de una residencia ronda los 2100 euros, lo que incentivó a la elección de los valores y rangos de renta anteriores, para así lograr el objetivo de morosidad nula.

Capa 4: Saturación Territorial (Peso: 15 %). Factor de corrección que capta el grado de saturación del mercado residencial mediante la relación entre la oferta y la demanda potencial a nivel provincial. Este indicador se define como el cociente entre el número total de plazas residenciales existentes en cada provincia y la población provincial de 80 o más años, considerada como población objetivo. De este modo, el factor aproxima qué proporción de dicha población se encuentra ya cubierta por la oferta disponible, permitiendo ajustar la predicción del algoritmo en función del nivel de saturación del mercado regional. Se normaliza a [0,1] e invierte (valores bajos de saturación = *score* altos).

Score SEED Final. La combinación ponderada de las tres capas produce el score SEED:

$$\text{SEED} = 0,45 * \text{Score_Demanda} + 0,40 * \text{Score_Renta} + 0,15 * \text{Score_Saturación}$$

Algoritmo de Clustering Espacial. El *score* SEED por sí solo no garantiza una distribución óptima. Es necesario evitar la canibalización: residencias demasiado cercanas compiten por los mismos clientes. Para ello, se implementa un algoritmo *greedy* iterativo con restricciones espaciales adaptativas.

Cálculo de Distancias Geográficas. Para cuantificar la separación entre secciones censales a partir de sus coordenadas geográficas (latitud, longitud), se utiliza la distancia ortodrómica mediante la *fórmula de Haversine*, adecuada para distancias sobre la superficie terrestre.

Sean (φ_1, λ_1) y (φ_2, λ_2) las latitudes y longitudes (en radianes) de dos secciones censales. Se define:

$$\Delta\varphi = \varphi_2 - \varphi_1, \quad \Delta\lambda = \lambda_2 - \lambda_1, \quad [1]$$

$$a = \sin^2\left(\frac{\Delta\varphi}{2}\right) + \cos(\varphi_1) \cos(\varphi_2) \sin^2\left(\frac{\Delta\lambda}{2}\right), \quad [2]$$

$$c = 2 \arctan 2(\sqrt{a}, \sqrt{1-a}), \quad [3]$$

y la distancia en kilómetros queda:

$$d_{\text{km}} = R c, \quad [4]$$

donde $R = 6371$ km es el radio medio de la Tierra.

Restricción Adaptativa de Distancia Mínima. Las distancias mínimas de separación entre ubicaciones se definen de forma adaptativa en función de la densidad poblacional de cada sección censal, con el objetivo de evitar la sobresaturación territorial y garantizar una distribución espacial equilibrada:

- Zonas de alta densidad ($> 5,000$ habitantes/km²): $d_{\text{mín}} = 1,5$ km
- Zonas de densidad media ($1,000$ – $5,000$ habitantes/km²): $d_{\text{mín}} = 2,5$ km
- Zonas de baja densidad (rurales, $< 1,000$ habitantes/km²): $d_{\text{mín}} = 5,0$ km

Cabe destacar que las distancias tenidas en cuenta mantienen la accesibilidad de trabajadores y familiares a la respectiva ubicación de la residencia.

Algorithm 1 Selección Espacial Adaptativa basada en el índice SEED

```
1: Ordenar todas las secciones censales por puntuación SEED descendente
2: Inicializar lista vacía  $\mathcal{S}$  (ubicaciones seleccionadas)
3: for cada sección  $s$  en el orden establecido do
4:   if  $|\mathcal{S}| = 1000$  then
5:     finalizar
6:   Determinar  $d_{\min}(s)$  en función de la densidad poblacional
7:   Calcular las distancias entre  $s$  y todas las ubicaciones en  $\mathcal{S}$ 
8:   if todas las distancias  $> d_{\min}(s)$  then
9:     Añadir  $s$  a  $\mathcal{S}$ 
10: return  $\mathcal{S}$  con ranking del 1 al 1000
```

Sistema de Visualización y Conclusiones

Mapas de Gradiente de Calor. Tras la ejecución del algoritmo se generan mapas interactivos HTML para cada observar el impacto de capa individual del algoritmo, utilizando la librería folium con plugin HeatMap:

- Mapa de Demanda Residencial: Visualiza el top 5 % de secciones por *score* de demanda.
- Mapa de Viabilidad Económica: Destaca el top 5 % de zonas con renta óptima (cerca a 72000 euros).
- Mapa de Saturación: Identifica oportunidades en zonas poco saturadas (top 5 %).
- Mapas de Resultados: Combinación ponderada de todas las capas antes y después de aplicar el algoritmo de *clustering*.

Mapa de las 1000 Ubicaciones Seleccionadas. El mapa final muestra las 1000 ubicaciones con marcadores circulares rojos. Cada marcador tiene en cuenta el ranking de la ubicación, el código de sección censal, la puntuación final, la provincia a la que pertenece, y las coordenadas geográficas exactas. Este mapa se apoya de su respectivo *dataframe* con el ranking de top 1000 ubicaciones situadas tras haber aplicado el *clustering*.

Mapa de TOP 50. Este mapa representa un modo interactivo y visual de localizar las 50 mejores residencias dentro del ranking. Como el anterior, se apoya también en su *dataframe*.

DataFrames. Además de los *dataframes* con los resultados óptimos y definitivos ya mencionados, por cada capa se extrae el ranking actualizado en un archivo de datos tras haber aplicado la función correspondiente de su celda. Por otro lado, también se genera un archivo de datos con el ranking global de las puntuaciones obtenidas de los aproximadamente 36.000 censos, sin aplicar *clustering* para conocer y aportar la puntuación inidividual de cada sección.

Análisis estadístico. El análisis del algoritmo SEED identifica a Galicia, Comunidad Valenciana, Andalucía, Cataluña y Madrid como las regiones con el mayor potencial estratégico para la apertura de nuevas residencias. La concentración del Top 1000 en estas zonas, con A Coruña (15,7 %), Pontevedra (9,2 %), Alicante (8,0 %), Madrid (5,7 %) y Barcelona (5,2 %) a la cabeza, no es casual; señala directamente los mercados donde la oportunidad de inversión es máxima debido a la combinación del potencial de la demanda y la capacidad económica de cada región. Además, cabe resaltar que la distribución planteada se mantiene coherente con el déficit habitacional identificado por la AEDGSS en dichas regiones y contribuye a paliar la falta de plazas en zonas especialmente afectadas, como Madrid y Andalucía.

Así pues, la fiabilidad del modelo se sustenta en una correlación de 0,882 respecto a la distribución real del sector al aislar el factor de saturación, demostrando un ajuste preciso a la realidad operativa del mercado.

A nivel macro, los censos en el Top 50 promedian una renta de 72.769,89 € y un F-of-M de 0,1194, cifras que se alinean casi perfectamente con el "punto ideal" definido en el modelo.

La eficacia del algoritmo se confirma también en el análisis de microlocalización. Por ejemplo, el caso de la sección censal 1503003001 en A Coruña, con el score SEED más alto (0,838), presenta una renta de 74.388 € y un F-of-M de 0,0979. Estos indicadores, sumados a una densidad de población óptima y baja saturación de la sección censal, demuestran que el algoritmo cumple su función de transformar datos demográficos complejos en recomendaciones estratégicas probadas

Vista al futuro. Nuestro algoritmo no pretende ser un modelo definitivo, sino una herramienta de apoyo a la toma de decisiones estratégicas que puede perfeccionarse con el tiempo; aunque los resultados son prometedores, existen aspectos susceptibles de mejora para aumentar su precisión y fiabilidad.

Su naturaleza ponderativa introduce cierto grado de subjetividad en la selección de variables y en la asignación de pesos, lo que, junto con la similitud de perfiles en la parte baja del ranking, deja margen de mejora para la diferenciación entre localizaciones cuando sus valores se alejan del perfil ideal. En esta línea, la incorporación de variables más específicas, como una mayor segmentación de perfiles de residente, diferencias de precios por región y entre oferta pública y privada, y el uso de datos más detallados a nivel local o censal, en lugar de indicadores amplios autonómicos

o provinciales, permitirían reducir ruido en el modelo y generar resultados más realistas y ajustados a la realidad territorial.

Considerando lo anteriormente mencionado, las mejoras planteadas no cuestionan la base de nuestro modelo, sino que refuerzan una dirección que ya ha mostrado ser coherente y prometedora. Al fin y al cabo, cualquier herramienta analítica útil debe aspirar a revisarse y perfeccionarse de forma continua conforme mejora la información disponible.