

Algoritmo L-SOMA

*Sistema de Localización Óptima para
Residencias de Mayores en España*

Caso práctico: Desarrolla un algoritmo de IA de venta
Fundación Innovación Bankinter: Akademia

Johan Manuel Orozco Mesa

Enero 2026

Índice

1. Resumen Ejecutivo	3
1.1. Resultados Principales: Escenarios de Absorción	3
2. Marco Teórico y Justificación Metodológica	3
2.1. Del Análisis Descriptivo al Espacio de Hilbert	3
2.2. El Vector Objetivo: La Pirámide Ideal	4
2.2.1. Parametrización del Vector Objetivo $ Q\rangle$	4
2.3. Métrica de Resonancia: Divergencia de Jensen-Shannon	5
2.3.1. Propiedades de la JSD	6
3. Fuentes de Datos y Tratamiento	6
3.1. Fuentes Primarias del INE	6
3.2. Proceso de Ingesta y Limpieza	6
3.2.1. Fase 1: Extracción de Datos Demográficos	6
3.2.2. Fase 2: Construcción de la Matriz P	7
3.2.3. Fase 3: Integración de Datos Económicos	8
3.3. Auditoría de Calidad de Datos	8
4. Metodología: El Algoritmo L-SOMA	8
4.1. Arquitectura del Sistema	8
4.2. Fase 1: Cálculo de Resonancia Demográfica	9
4.3. Fase 2: Filtro Económico (Solvencia)	10
4.4. Fase 3: Feature Engineering de Disparadores Sociales	10
4.4.1. Variable A: Ratio de Presión de Cuidados (Π)	10
4.4.2. Variable B: Factor de Burnout Normalizado (β)	11
4.5. Fase 4: Función de Score Global	11
4.6. Fase 5: Clustering Espacial (DBSCAN)	11
5. Resultados: Análisis de Clustering	12
5.1. Estadísticas Globales de Clustering	12
5.2. Top 10 Clusters por Potencia Total	13
5.3. Visualización Geográfica de Clusters	14
6. Resultados: Análisis de Masa Crítica y Viabilidad	14
6.1. Concepto de Masa Crítica	14
6.2. Cálculo de Demanda Efectiva por Sección	15
6.3. Modelo de Viabilidad Operativa	15
6.4. Escenarios de Penetración de Mercado	15
6.5. Distribución Geográfica de Clusters Viables	17
7. La Frontera de Rentabilidad: Algoritmo de Expansión Adaptativa	17
7.1. Motivación	17
7.2. Diseño del Algoritmo de Expansión	18
7.2.1. Jerarquía de Riesgo	18
7.2.2. Criterio de Selección Inteligente	18
7.3. Resultados de la Expansión	18

7.3.1. Parámetros Finales	19
7.4. Comparativa: Modelo Prime vs Modelo Expandido	19
7.5. Distribución Geográfica de Clusters Expandidos	20
7.6. Interpretación del Gap Restante	21
7.7. Análisis de Rendimientos Decrecientes	21
7.7.1. Punto de Inflexión: Iteración 40	21
7.7.2. Conclusión sobre la Capacidad Real del Mercado	22
7.8. Implicaciones Estratégicas	22
8. Validación de Competencia: Análisis de Oferta Existente	23
8.1. Motivación del Análisis	23
8.2. Metodología	23
8.2.1. Parámetros del Análisis	23
8.2.2. Cálculo del Índice de Saturación	23
8.2.3. Clasificación Estratégica	23
8.3. Resultados	23
8.4. Los 3 Blue Oceans Identificados	24
8.4.1. Análisis de los Blue Oceans	25
8.5. Limitaciones del Análisis de Competencia	25
9. Discusión	25
9.1. Síntesis de Hallazgos Cuantitativos	25
9.2. Concentración Geográfica de la Demanda	26
9.3. Análisis del Gap Estratégico	26
9.4. Rendimientos Decrecientes y Punto Óptimo	27
9.5. Saturación del Mercado: Análisis de Competencia	27
9.6. Limitaciones del Estudio	27
10. Conclusiones	28
10.1. Validación Metodológica	28
10.2. Hallazgos Estratégicos	28
10.3. Recomendación Estratégica Final	29
A. Anexo A: Datos de Entrada	30
B. Anexo B: Pipeline de Procesamiento	30
C. Anexo C: Parámetros del Modelo	31
C.1. Parámetros Fijos (No Modificados)	31
C.2. Parámetros Variables (Escenarios)	31
C.3. Código: Divergencia Jensen-Shannon	31
C.4. Código: Clustering DBSCAN Geodésico	31
D. Anexo D: Validación de Competencia	32
D.1. Configuración de Google Places API	32
D.2. Filtros de Calidad Aplicados	32
D.3. Archivos Generados	32

1. Resumen Ejecutivo

El presente informe documenta el desarrollo del algoritmo **L-SOMA** (*Location-based System for Optimal Market Allocation*), un sistema de soporte a la decisión diseñado para identificar las localizaciones óptimas para la apertura de residencias de mayores en España. El análisis se fundamenta en el estudio previo “*Disparadores de Ingreso a Residencias Privadas*”, que identifica los factores clave de demanda.

El objetivo del proyecto es evaluar la viabilidad de un plan de expansión de **1.000 nuevas residencias** de **100 camas** cada una, con un coste de plaza de **2.100€ mensuales**, garantizando:

- Una ocupación media del 85 % para asegurar el retorno de inversión.
- Minimización del riesgo de morosidad mediante filtros de solvencia.
- Identificación de “océanos azules” con alta demanda demográfica y baja competencia.

1.1. Resultados Principales: Escenarios de Absorción

Tras aplicar el algoritmo sobre las 32.910 secciones censales, se ha calculado la capacidad de absorción del mercado. Para ello, se define la variable **Cuota de Mercado** como el porcentaje de la población objetivo (mujeres mayores de 80 años solventes) que efectivamente optaría por ingresar en nuestras residencias frente a alternativas públicas, competencia o cuidado domiciliario.

Considerando que se trata de un nuevo operador en el mercado (entrante), se modelan cuatro escenarios de sensibilidad conservadores:

Cuadro 1: Viabilidad del Plan de Expansión según Tasa de Captura*

Escenario (Cuota)	Clusters Viables	Residencias	Gap vs 1.000
1,5 % (Muy Conservador)	6	17	-983
2,0 % (Conservador)	12	29	-971
3,0 % (Base)	19	50	-950
5,0 % (Agresivo)	25	91	-909

*Escenarios bajo restricciones de calidad “Prime” (Renta >30k€, Camas ≥ 85). El Modelo Expandido (Sección 7) relaja estos parámetros alcanzando 359 residencias.

Conclusión estratégica: El mercado español soporta entre 27 y 117 residencias nuevas mediante construcción orgánica. Para alcanzar el objetivo de 1.000 unidades, es imperativa una estrategia complementaria de fusiones y adquisiciones (M&A).

2. Marco Teórico y Justificación Metodológica

2.1. Del Análisis Descriptivo al Espacio de Hilbert

El enfoque tradicional para localización de negocios se basa en estadísticas descriptivas simples (media de edad, densidad de población). Sin embargo, este enfoque pierde información estructural crítica.

Definición 1 (Pérdida de Información Escalar). *Si describimos una población únicamente con su edad media (un escalar), estamos proyectando un sistema complejo en una sola dimensión. Ejemplo: un grupo de niños de 5 años y ancianos de 95 años tiene la misma media de 50 años que un grupo homogéneo de adultos de 50 años. Para un negocio de residencias, estos escenarios son opuestos.*

Para superar esta limitación, adoptamos un enfoque de **física de sistemas**, tratando cada sección censal como un **vector de estado** en un espacio de Hilbert n -dimensional.

Definición 2 (Vector de Estado y Normalización L_1). *Sea $\vec{C}_i \in \mathbb{N}^{42}$ el vector de conteos absolutos (datos crudos) de la sección censal i , donde cada componente $n_{s,r}$ representa el número de personas del sexo s y rango de edad r .*

*La magnitud de \vec{C}_i depende del tamaño poblacional de la sección, lo que introduce un sesgo de escala que impide comparaciones estructurales directas. Para resolver esto, definimos el **Vector de Estado Normalizado** $|P_i\rangle$ aplicando una proyección al simplex estándar de probabilidad (normalización L_1):*

$$|P_i\rangle = \frac{\vec{C}_i}{\|\vec{C}_i\|_1} = \frac{1}{N_i}(n_{H,0-4}, \dots, n_{M,100+}) \quad (1)$$

donde $N_i = \sum_{s,r} n_{s,r}$ es la población total de la sección. Esta operación transforma los conteos brutos en una Función de Masa de Probabilidad (PMF), garantizando la condición de cierre:

$$\langle P_i | \mathbf{1} \rangle = \sum_{s,r} p_{s,r} = 1 \quad (2)$$

De este modo, desacoplamos la “forma” de la pirámide demográfica (estructura) de su “volumen” (tamaño), permitiendo medir la distancia termodinámica entre poblaciones de magnitudes dispares.

2.2. El Vector Objetivo: La Pirámide Ideal

Basándonos en el estudio sociológico (ver Anexo A), definimos el vector objetivo $|Q\rangle$ que representa la distribución demográfica “ideal” para el negocio de residencias.

2.2.1. Parametrización del Vector Objetivo $|Q\rangle$

Dado que el estudio de referencia establece que la **edad media de ingreso se ha retrasado hasta los 82 años**, la distribución de pesos del vector $|Q\rangle$ no puede ser lineal. Debe modelarse como una curva que penalice las edades tempranas y maximice la señal en la “Cuarta Edad”.

Los pesos $w_{s,r}$ se han calibrado para satisfacer la restricción de edad media del estudio:

1. **Peso de Continuidad (60-74 años) = 0,5:** Se asigna un valor reducido para capturar la **Tasa de Reposición**. Una zona con alta densidad de 85+ pero vacío en la franja 60-74 es un “mercado en extinción”; este peso premia la sostenibilidad del flujo de clientes a medio plazo (horizonte 10 años).
2. **Peso Base (75-79 años) = 1:** Se establece como unidad de referencia. El ingreso en esta franja es bajo porque las familias aún intentan mantener el cuidado en el hogar.

3. **Peso Máximo (85+ años) = 5:** Para obtener una media de entrada de 82 años, la probabilidad de ingreso en el segmento 85+ debe ser significativamente superior (aprox. 5 veces) a la del segmento 75-79. Esto refleja la situación de “fragilidad extrema” e “inmanejabilidad” descrita en el informe.
4. **Factor de Soledad ($\gamma = 1, 3$):** El estudio destaca que casi 2 millones de mayores viven solos y que la soledad es un factor crítico. Dado que demográficamente la soledad en la cuarta edad tiene mayor prevalencia femenina, aplicamos un factor de corrección de 1,3 a los segmentos femeninos para capturar este riesgo.

Cuadro 2: Matriz de Pesos calibrada según Edad Media de Ingreso (82 años)

Rango	Peso H (w_H)	Peso M (w_M)	Fundamento en el Estudio
60-74	0,50	0,65	Demanda Latente (Sostenibilidad futura)
75-79	1,00	1,30	Resistencia familiar al ingreso
80-84	4,00	5,20	Entorno de la media (82 años)
85+	5,00	6,50	Situación “inmanejable” / Fragilidad

Nota sobre la Normalización: La suma bruta de estos pesos ($W_{total} = \sum w_{s,r}$) no es unitaria. Para obtener el vector de estado físico $|Q\rangle$ que cumpla las condiciones de un espacio de probabilidad ($\sum Q = 1$), se aplica una normalización L_1 final:

$$|Q\rangle = \left(\frac{w_{H,0-4}}{W_{total}}, \dots, \frac{w_{M,100+}}{W_{total}} \right) \quad (3)$$

De esta forma, garantizamos que $\langle Q|Q\rangle_1 = 1$, permitiendo la comparación directa con los vectores de población $|P_i\rangle$.

2.3. Métrica de Resonancia: Divergencia de Jensen-Shannon

Para cuantificar la “similitud” entre una sección censal $|P_i\rangle$ y nuestro vector objetivo $|Q\rangle$, empleamos la **Divergencia de Jensen-Shannon (JSD)**.

Definición 3 (Divergencia de Jensen-Shannon). Sean P y Q dos distribuciones de probabilidad sobre el mismo espacio muestral. La divergencia JSD se define como:

$$JSD(P\|Q) = \frac{1}{2}D_{KL}(P\|M) + \frac{1}{2}D_{KL}(Q\|M) \quad (4)$$

donde $M = \frac{1}{2}(P + Q)$ es la mezcla de ambas distribuciones, y D_{KL} es la divergencia de Kullback-Leibler:

$$D_{KL}(P\|Q) = \sum_x P(x) \log \frac{P(x)}{Q(x)} \quad (5)$$

Interpretación Física de la Divergencia KL: La Divergencia de Kullback-Leibler, también conocida como **Entropía Relativa**, mide la cantidad de información perdida (en bits, si el logaritmo es base 2) cuando se utiliza la distribución Q para aproximar la distribución verdadera P . En términos termodinámicos, cuantifica la “sorpresa” o el coste entrópico de asumir que una población sigue el patrón Q cuando en realidad se comporta como P . Cabe destacar que D_{KL} no es una distancia métrica real, ya que carece de simetría ($D_{KL}(P\|Q) \neq D_{KL}(Q\|P)$) y no satisface la desigualdad triangular, motivo por el cual recurrimos a la JSD para simetrizar la medición.

2.3.1. Propiedades de la JSD

La elección de JSD sobre otras métricas (distancia euclídea, divergencia KL) se justifica por:

1. **Simetría:** $JSD(P\|Q) = JSD(Q\|P)$.
2. **Acotación:** $0 \leq JSD \leq 1$ (usando logaritmo base 2).
3. **Robustez:** No explota cuando $P(x) > 0$ y $Q(x) = 0$.
4. **Interpretación:** Mide la “información compartida” entre distribuciones.

Definimos la **Resonancia Demográfica** como:

$$\mathcal{R}_i = 1 - JSD(P_i\|Q) \quad (6)$$

Un valor de \mathcal{R}_i cercano a 1 indica alta similitud con el perfil demográfico ideal; cercano a 0 indica disonancia total.

3. Fuentes de Datos y Tratamiento

3.1. Fuentes Primarias del INE

El proyecto utiliza exclusivamente datos públicos del Instituto Nacional de Estadística (INE):

1. **Censo de Población 2021-2025:** Distribución por sexo y edad quinquenal para las 32.910 secciones censales de España.
2. **Atlas de Distribución de Renta de los Hogares 2023:** Renta neta media por hogar a nivel de sección censal.
3. **Cartografía Censal 2024:** Shapefiles con los polígonos georreferenciados de cada sección censal (CUSEC).

3.2. Proceso de Ingesta y Limpieza

3.2.1. Fase 1: Extracción de Datos Demográficos

Los datos del censo se obtienen en formato CSV con separador tabulador, conteniendo las siguientes columnas:

- **Secciones:** Código CUSEC + nombre descriptivo
- **Sexo:** “Hombres”, “Mujeres” o “Total”
- **Edad:** Rango quinquenal (“De 0 a 4 años”, etc.) o “Todas las edades”
- **Periodo:** Año de referencia
- **Total:** Conteo de población

Filtros aplicados:

1. Exclusión de filas con Sexo = ‘Total’ (evita duplicación de datos).
2. Exclusión de filas con Edad = ‘Todas las edades’ (agregados que duplicarían los rangos específicos).
3. Selección del año más reciente disponible (2025).

Transformación de formato: El formato de edad del INE (“De 0 a 4 años”) se transforma al formato normalizado (“0-4”) mediante:

```
df['Edad_Clean'] = (df['Edad']
    .str.replace('años', ''))
    .str.replace('De ', '')
    .str.replace('a ', '-')
```

3.2.2. Fase 2: Construcción de la Matriz P

La matriz de población **P** tiene dimensiones $32,910 \times 42$, donde cada fila es una sección censal y cada columna un bin edad-sexo (H_0-4, H_5-9, ..., M_100+).

Para convertir conteos absolutos a probabilidades:

$$p_{i,j} = \frac{n_{i,j}}{\sum_j n_{i,j}} \quad (7)$$

donde $n_{i,j}$ es el conteo absoluto de la sección i en el bin j .

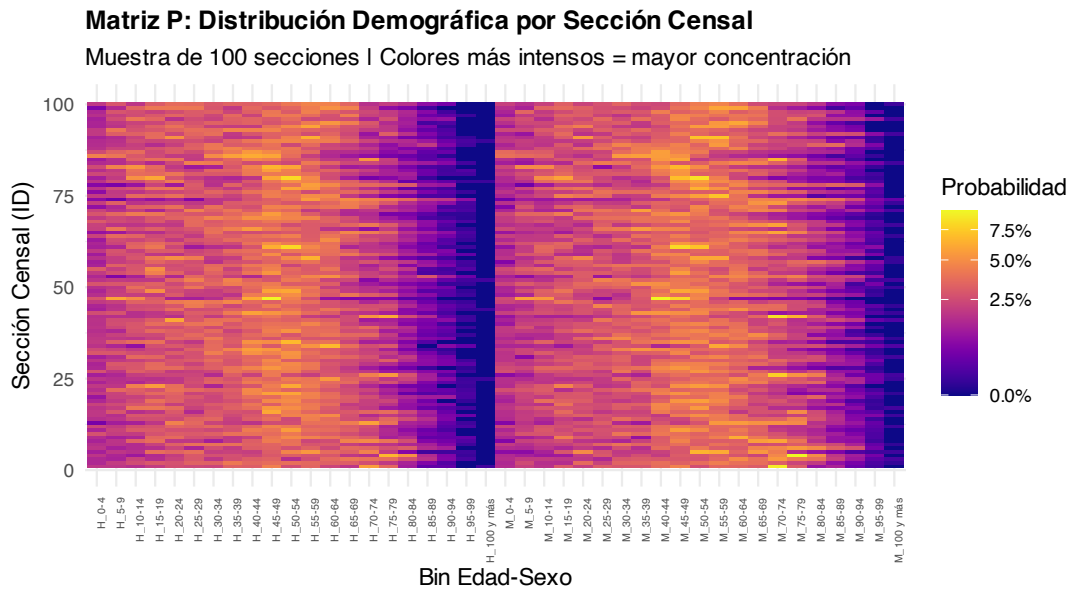


Figura 1: Visualización de la Matriz P para una muestra de 100 secciones aleatorias. Cada fila representa una sección censal; cada columna un bin edad-sexo. Colores más intensos indican mayor concentración de población.

Interpretación de la Matriz: La visualización revela una **dualidad estructural**. En los rangos centrales (adultos de 30–65 años), se observa una marcada *homogeneidad vertical* (bandas continuas naranjas/amarillas), indicando que la base poblacional es consistente en todo el territorio.

Sin embargo, en el segmento objetivo (M_{80+}), esta uniformidad se rompe. Las bandas verticales se fragmentan en **patrones discretos e intermitentes**. No todas las secciones censales conservan la “cola” de la distribución demográfica con la misma intensidad. Esta *variabilidad específica en la cuarta edad* —donde algunas filas muestran alta densidad (amarillo) y sus vecinas nula (morado)— es la que permite al algoritmo discriminar oportunidades de inversión frente al ruido de fondo general.

3.2.3. Fase 3: Integración de Datos Económicos

Los datos de renta se cruzan mediante el código CUSEC (primeros 10 dígitos de la columna Secciones):

```
df_rank['CUSEC'] = df_rank['Seccion'].str.split('_').str[0]
df_final = pd.merge(df_rank, df_renta[['CUSEC', 'Renta_Hogar']],
                    on='CUSEC', how='left')
```

Tratamiento de valores faltantes: Las secciones sin dato de renta (aproximadamente 5 %) reciben un valor de 0, lo que las penaliza en el score final.

3.3. Auditoría de Calidad de Datos

Antes de proceder al análisis, se ejecuta una auditoría automática que verifica:

1. **Estructura:** Las 42 columnas de la matriz P coinciden con el vector Q.
2. **Masa:** La población total representada está en el rango esperado (40-50 millones).
3. **Normalización:** Cada fila suma exactamente 1.0 (\pm error de punto flotante).

4. Metodología: El Algoritmo L-SOMA

4.1. Arquitectura del Sistema

El algoritmo L-SOMA se estructura en cinco capas de procesamiento:

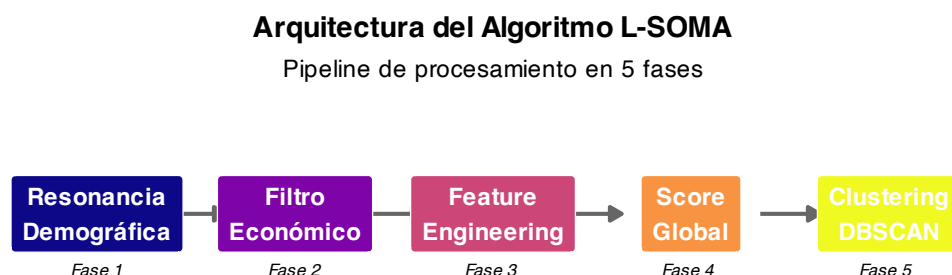


Figura 2: Arquitectura del algoritmo L-SOMA.

4.2. Fase 1: Cálculo de Resonancia Demográfica

Para cada sección censal i , calculamos la distancia Jensen-Shannon respecto al vector objetivo:

$$d_i = JSD(P_i \| Q) \quad (8)$$

La implementación utiliza `scipy.spatial.distance.jensenshannon`:

```
from scipy.spatial.distance import jensenshannon
distancias = np.apply_along_axis(
    lambda row: jensenshannon(row, vector_Q),
    axis=1, arr=matriz_P)
resonancias = 1.0 - distancias
```

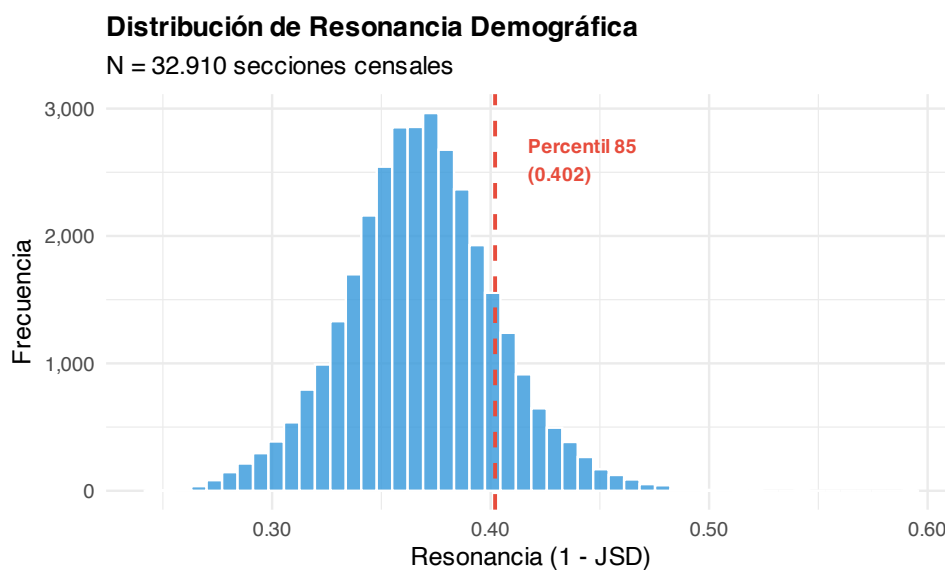


Figura 3: Distribución de la Resonancia Demográfica \mathcal{R} para las 32.910 secciones censales. La línea vertical indica el umbral del percentil 85.

Análisis de Selectividad: El histograma muestra que la Resonancia Demográfica sigue una distribución cuasi-normal centrada en $\mu \approx 0,38$. Es crucial notar que ninguna sección alcanza el valor teórico de 1,0; los mejores candidatos saturan en torno a 0,50, lo que evidencia la complejidad del tejido social real frente al modelo ideal.

Criterio de Corte (Percentil 85): Para optimizar los recursos computacionales y enfocar la inversión, establecemos un umbral de corte estricto en el percentil 85 ($R > 0,402$). Esto actúa como un filtro de calidad que descarta automáticamente el 85% del territorio nacional, reduciendo el espacio de búsqueda de 32,910 a las **4.937 secciones más prometedoras** antes de aplicar los criterios económicos. El umbral de 85 camas corresponde a la capacidad mínima considerada viable operativamente por los principales operadores del sector, superando en un 15% la media nacional de 74,8 plazas/centro (IMSERSO, 2023).

4.3. Fase 2: Filtro Económico (Solvencia)

El modelo incorpora una función de transferencia no lineal $\Phi(r)$ que modula la resonancia demográfica según la viabilidad financiera de la zona. Se utiliza la Renta Neta Media por Hogar (r) como variable independiente.

$$\Phi_{econ}(r) = \begin{cases} 0,4 & \text{si } r < 30,000\text{€} \\ \frac{65,000}{r} & \text{si } r > 65,000\text{€} \\ 1,0 & \text{en otro caso} \end{cases} \quad (9)$$

Fundamentación de los Parámetros:

1. **Coefficiente de Fricción** ($\Phi = 0,4$ para $r < 30k$): Evitamos un filtrado binario (0) para capturar el fenómeno *Asset Rich, Cash Poor*. Aunque la renta de flujo (pensión) sea insuficiente, gran parte del target posee vivienda en propiedad susceptible de licuación patrimonial (venta/alquiler) para financiar la plaza. El factor 0,4 modela la **mayor resistencia a la conversión** (fricción de venta) en este segmento, penalizando el score sin descartar la demanda latente.
2. **Decaimiento Hiperbólico** ($\Phi \propto 1/r$ para $r > 65k$): Se aplica una penalización inversa para rentas altas. Aunque la capacidad de pago es total, estas zonas presentan un coste de suelo (repercusión) que dispara el **CAPEX** (Inversión de Capital Inicial). Dado que el precio de la plaza tiene un techo de mercado ($\approx 2,500\text{€}$), un suelo excesivamente caro erosiona la Tasa Interna de Retorno (TIR). Esta función prioriza zonas de clase media donde el binomio coste-suelo/precio-venta es óptimo. Además, existe un desajuste entre la propuesta de valor (2.100€/mes) y las expectativas del segmento *High Net Worth*, que demanda servicios de lujo (mayor de 4.000€) fuera del alcance de este proyecto. Evitamos así la canibalización por competidores del segmento *Luxury*.

4.4. Fase 3: Feature Engineering de Disparadores Sociales

4.4.1. Variable A: Ratio de Presión de Cuidados (Π)

Basándonos en el concepto de *Caregiver Burnout*, calculamos el ratio de presión de cuidados para cada sección censal i :

$$\Pi_i = \frac{\text{Mujeres}_{80+,i}}{\text{Mujeres}_{45-64,i} + \epsilon} \quad (10)$$

donde $\epsilon = 0,001$ evita división por cero.

Interpretación: Este ratio mide la carga potencial de cuidados sobre la cohorte de mujeres de mediana edad (principales cuidadoras informales según el INE).

- $\Pi_i \gg 1$: Alta presión asistencial; déficit de cuidadoras familiares en la sección.
- $\Pi_i \ll 1$: Baja presión; mayor capacidad de soporte familiar en el hogar.

4.4.2. Variable B: Factor de Burnout Normalizado (β)

Para asegurar la estabilidad del modelo y evitar que valores extremos (outliers) en secciones muy envejecidas distorsionen el Score Global, aplicamos una normalización centrada en la media nacional $\bar{\Pi}$:

$$\beta_i = \text{clip} \left(\frac{\Pi_i}{\bar{\Pi}}, 0,5, 1,5 \right) \quad (11)$$

El uso de la función *clip* limita el impacto de la variable a un rango de $[0,5, 1,5]$, evitando que un solo factor anule o dispare artificialmente el resultado final.

4.5. Fase 4: Función de Score Global

El score L-SOMA integra todas las variables mediante:

$$\mathcal{S}_i = \mathcal{R}_i \cdot \Phi_{econ}(r_i) \cdot \beta_i \quad (12)$$

Esta formulación multiplicativa garantiza que:

- Una sección con resonancia alta pero renta baja obtiene score bajo (filtro económico).
- Una sección con resonancia alta y renta adecuada pero baja presión de cuidados obtiene score moderado.
- Solo las secciones con **todos** los factores favorables alcanzan scores máximos.

4.6. Fase 5: Clustering Espacial (DBSCAN)

Las secciones individuales no son unidades de negocio viables. Una residencia de 100 camas requiere captar clientes de múltiples secciones adyacentes. Aplicamos DBSCAN sobre los **centroides geográficos** de las secciones filtradas ($\mathcal{S}_i > \text{Umbral}$) para identificar “manchas” de alta densidad.

Definición 4 (DBSCAN con Métrica Haversine). *DBSCAN agrupa puntos basándose en la densidad espacial, robusto frente al ruido.*

- **Input:** Coordenadas (*Lat, Lon*) de las secciones con alto Score.
- **Distancia:** Fórmula de Haversine (distancia geodésica sobre la esfera terrestre).

Hiperparámetros seleccionados:

- $\epsilon = 1,5$ km (0,000235 rad): Define el radio máximo de captación peatonal/local.
- $\text{minPts} = 3$: Mínimo de secciones necesarias para formar un núcleo. Esto garantiza una masa crítica poblacional aproximada de 4.500 habitantes por cluster, reduciendo el riesgo de micro-mercados.

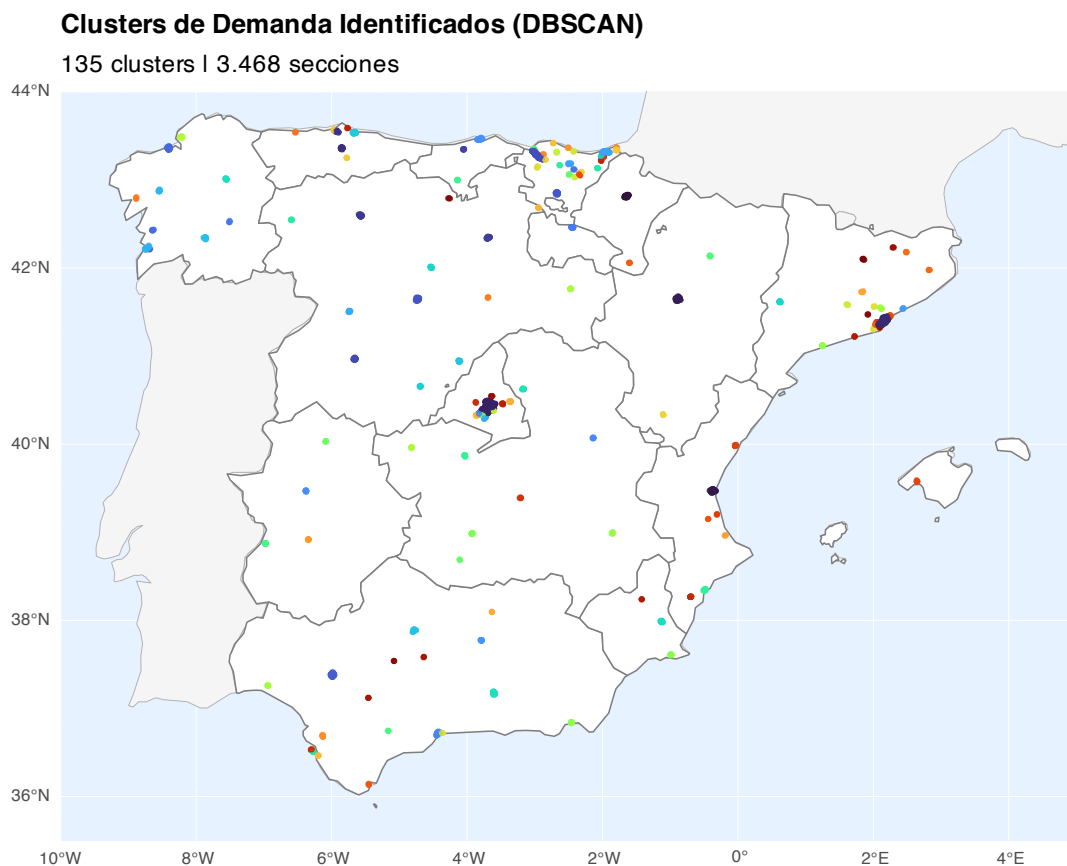


Figura 4: Distribución geográfica de los clusters detectados por DBSCAN. Cada color representa un cluster diferente.

5. Resultados: Análisis de Clustering

5.1. Estadísticas Globales de Clustering

Tras aplicar DBSCAN sobre las secciones del percentil 85+ (Top 15 % por Score Global), se obtienen:

Cuadro 3: Resumen del Clustering Espacial

Métrica	Valor
Secciones analizadas (Top 15 %)	3.468
Clusters detectados	135
Secciones en clusters viables	2.624
Secciones descartadas (ruido/aisladas)	844

5.2. Top 10 Clusters por Potencia Total

La “potencia” de un cluster se define como la suma de los scores individuales de sus secciones:

$$\text{Potencia}_c = \sum_{i \in c} \mathcal{S}_i \quad (13)$$

Cuadro 4: Top 10 Clusters de Mayor Potencia

Rank	Zona Representativa	Secciones	Renta Media	Potencia
1	Madrid	722	47.466€	421,7
2	Barcelona	323	45.313€	185,5
3	Bilbao	209	43.073€	124,3
4	Zaragoza	141	42.624€	83,1
5	Valladolid	98	40.330€	59,5
6	Donostia/San Sebastián	83	48.727€	48,8
7	Sevilla	82	48.743€	47,6
8	Pamplona/Iruña	74	46.973€	44,7
9	València	74	43.915€	42,3
10	Gijón	70	37.825€	41,1

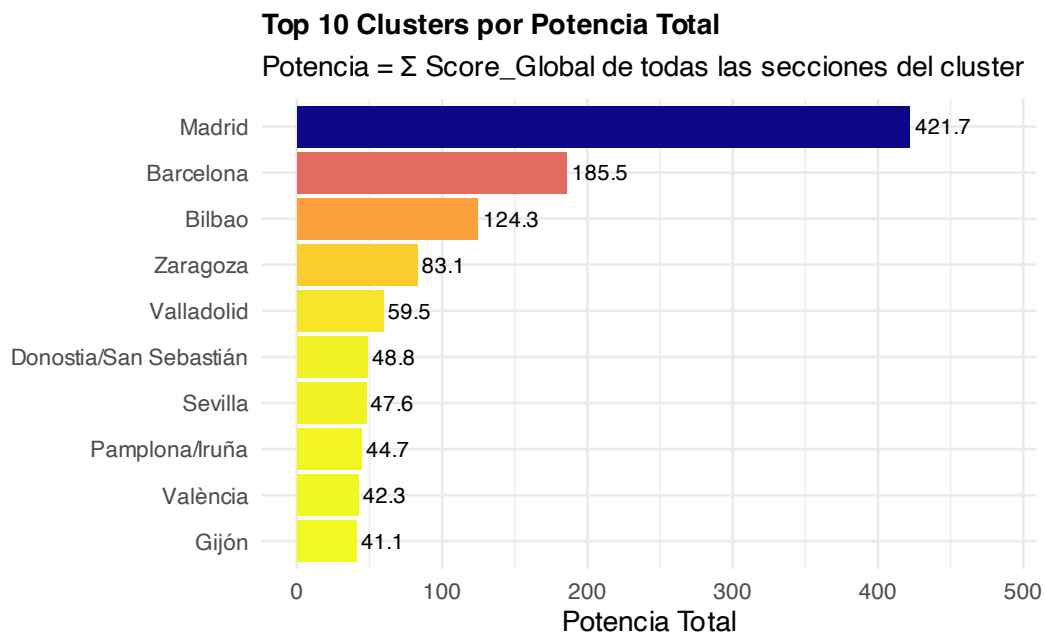


Figura 5: Comparativa de potencia de los Top 10 clusters identificados.

5.3. Visualización Geográfica de Clusters

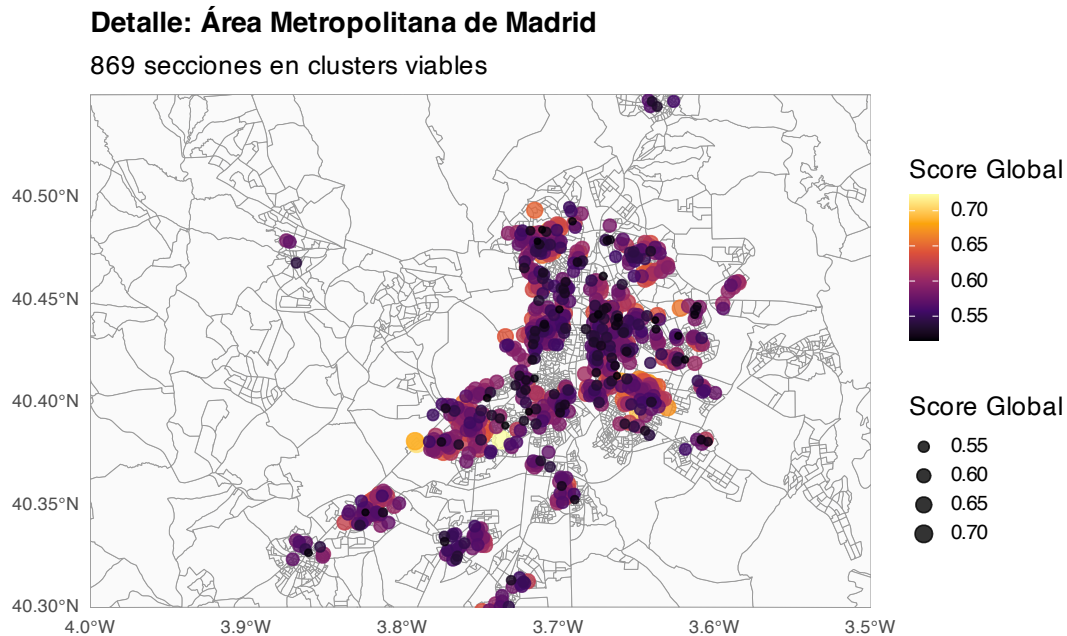


Figura 6: Detalle de clusters en el área metropolitana de Madrid. El color indica el Cluster ID; la intensidad representa el Score Global.

6. Resultados: Análisis de Masa Crítica y Viabilidad

6.1. Concepto de Masa Crítica

Definimos la **Masa Crítica** como el volumen mínimo de demanda latente necesario para justificar la inversión. En el sector de residencias, la viabilidad financiera no depende solo de la calidad demográfica (porcentajes), sino de la **escala absoluta**.

La necesidad de alcanzar esta masa crítica responde a la estructura de costes operativos (OPEX) del negocio:

- **Economías de Escala:** Una residencia estándar requiere una plantilla fija mínima (médico, enfermería 24h, cocina, limpieza) independientemente de la ocupación.
- **Umbral de Rentabilidad:** Por debajo de 80-90 camas, el coste fijo por residente se dispara, erosionando el margen EBITDA.
- **Mitigación de Riesgo:** Un *pool* de demanda amplio es necesario para amortiguar la rotación natural de plazas (fallecimientos) sin que la ocupación caiga por debajo del 90 %.

Por tanto, el algoritmo no busca solo “donde hay ancianos”, sino “donde hay *suficientes* ancianos” para llenar un activo de escala industrial.

6.2. Cálculo de Demanda Efectiva por Sección

Para cuantificar esta masa con precisión censal, el modelo calcula la población objetivo exacta (T_i) para cada sección i , integrando directamente las probabilidades de la Matriz P:

$$T_i = P_{i,\text{total}} \cdot \sum_{r \in \text{Target}} p_{i,r} \quad (14)$$

donde:

- $P_{i,\text{total}}$: Población total registrada en la sección censal.
- $p_{i,r}$: Probabilidad de pertenencia al grupo demográfico objetivo (Mujeres > 80 años) extraída de la matriz de distribución demográfica.

El análisis computacional arroja una media nacional de **70,1 mujeres target por sección censal**, dato que sirve de base para la agregación por clusters.

6.3. Modelo de Viabilidad Operativa

Una vez identificados los clusters espaciales, evaluamos su viabilidad transformando la demanda demográfica en capacidad de absorción de camas.

La **Capacidad Teórica** (C_c) de un cluster c se define como:

$$C_c = \sum_{i \in c} T_i \cdot \phi \quad (15)$$

donde ϕ es la cuota de mercado (*Market Share*) estimada, que representa el porcentaje de la población target que efectivamente optará por el servicio residencial privado.

Se establece una función de decisión binaria basada en el umbral de eficiencia operativa:

$$\text{Viable}_c = \begin{cases} 1(\text{Aprobado}) & \text{si } C_c \geq 85 \text{ camas} \\ 0(\text{Descartado}) & \text{en otro caso} \end{cases} \quad (16)$$

Este corte duro (*hard cut*) de 85 camas asegura que solo se seleccionen ubicaciones capaces de sostener los costes fijos estructurales descritos anteriormente. El umbral corresponde a la capacidad mínima considerada viable operativamente por los principales operadores del sector, superando en un 15 % la media nacional de 74,8 plazas/centro [16].

6.4. Escenarios de Penetración de Mercado

Dada la variabilidad inherente a la tasa de conversión comercial (ϕ), se presenta un análisis de sensibilidad que evalúa la robustez de los clusters ante diferentes escenarios.

Cuadro 5: Matriz de Viabilidad según Cuota de Mercado (ϕ)

Escenario	Share (ϕ)	Clusters Viables	Camas Potenciales	Gap vs Objetivo
Muy Conservador	1,5 %	6	17	-983
Conservador	2,0 %	12	29	-971
Base	3,0 %	19	50	-950
Agresivo	5,0 %	25	91	-909

Impacto del Market Share en la Viabilidad**Análisis de Sensibilidad**

Residencias viables según cuota de mercado

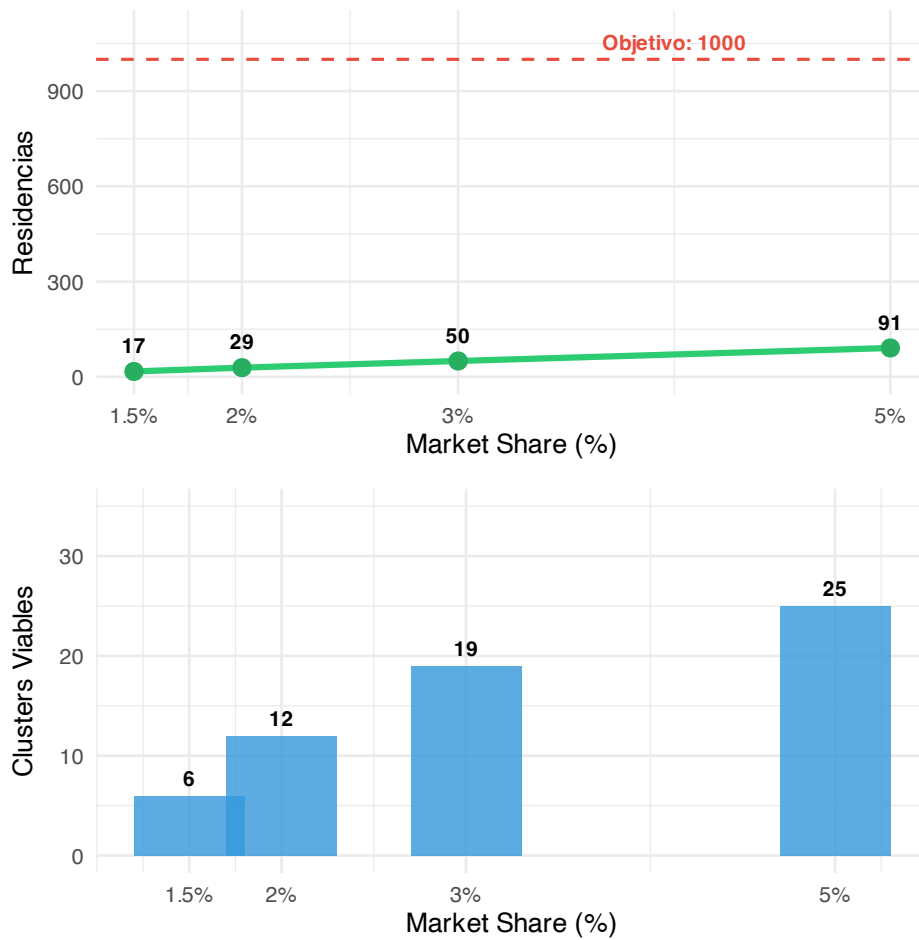


Figura 7: Curva de sensibilidad: Número de residencias operativamente viables en función de la cuota de mercado proyectada.

6.5. Distribución Geográfica de Clusters Viables

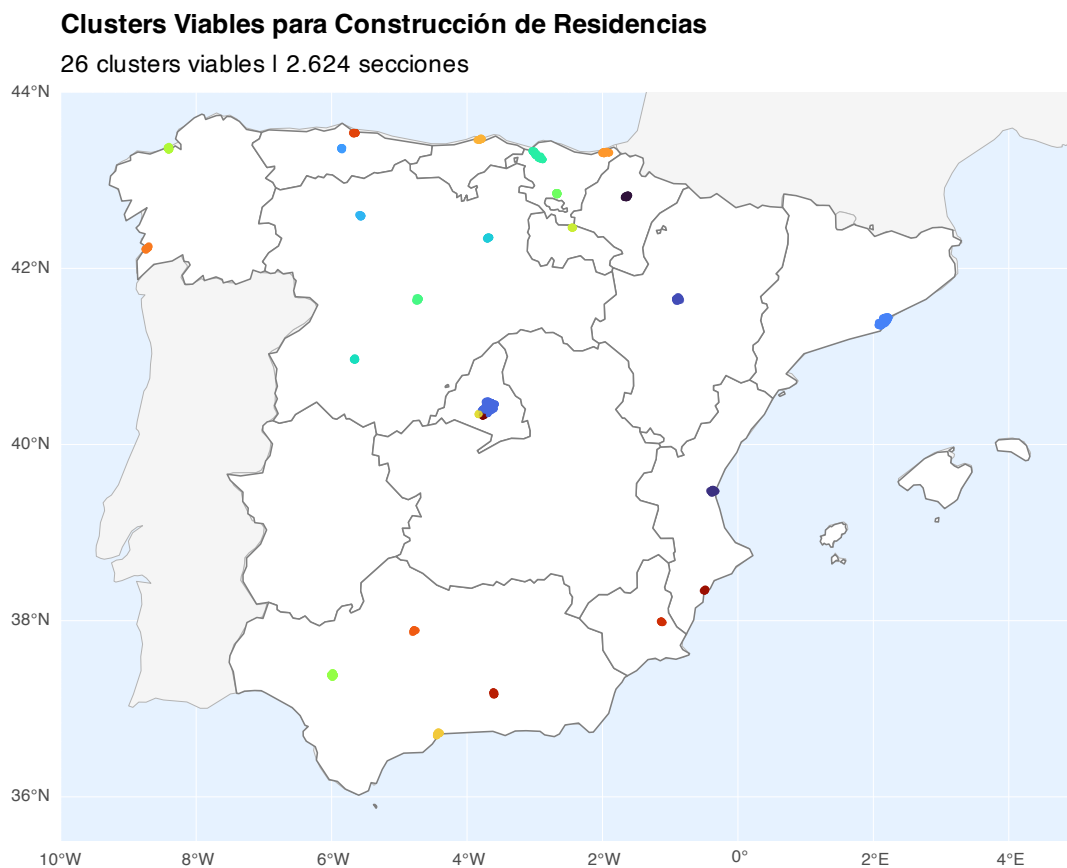


Figura 8: Distribución geográfica de los clusters viables para construcción de residencias. Cada color representa un cluster diferente con capacidad ≥ 85 camas.

7. La Frontera de Rentabilidad: Algoritmo de Expansión Adaptativa

7.1. Motivación

Los resultados del modelo Prime demuestran que, bajo restricciones estrictas de calidad demográfica y viabilidad operativa, el mercado español soporta únicamente **64 residencias** potenciales. Esta cifra dista significativamente del objetivo estratégico de 1.000 unidades.

Surge entonces la pregunta crítica: **¿Cuántas residencias adicionales podrían ubicarse si relajamos progresivamente las restricciones del modelo?** Para responderla, desarrollamos un algoritmo de *Descenso de Riesgo* que explora sistemáticamente el espacio de parámetros.

7.2. Diseño del Algoritmo de Expansión

El algoritmo parte de los parámetros “Prime” (óptimos) y entra en un bucle iterativo que relaja variables de forma inteligente, priorizando aquellas que ofrecen **mayor ganancia marginal con menor deterioro del modelo**.

7.2.1. Jerarquía de Riesgo

Se definen cuatro niveles de riesgo ordenados de menor a mayor impacto en la calidad del modelo:

Cuadro 6: Jerarquía de Variables para Relajación

Nivel	Variable	Inicio	Límite	Step
1 (Bajo)	Percentil de Score	85	60	-1
2 (Medio)	Market Share (ϕ)	3,0 %	6,0 %	+0,2 %
3 (Alto)	Penalización Renta	0,40	0,70	+0,05
4 (Crítico)	Camas Mínimas	85	60	-1

7.2.2. Criterio de Selección Inteligente

En cada iteración, el algoritmo **no desciende secuencialmente** por los niveles, sino que evalúa todas las opciones disponibles y selecciona aquella que maximiza la ganancia ajustada por riesgo:

$$\text{Ganancia}_{\text{ajustada}} = \frac{\Delta \text{Residencias}}{\text{Nivel de Riesgo}} \quad (17)$$

Este enfoque garantiza que se agote el potencial de los parámetros de bajo riesgo antes de recurrir a relajaciones más agresivas.

7.3. Resultados de la Expansión

El algoritmo ejecutó **71 iteraciones** hasta alcanzar el límite de expansión en todos los parámetros. La Figura 9 muestra la curva de convergencia.

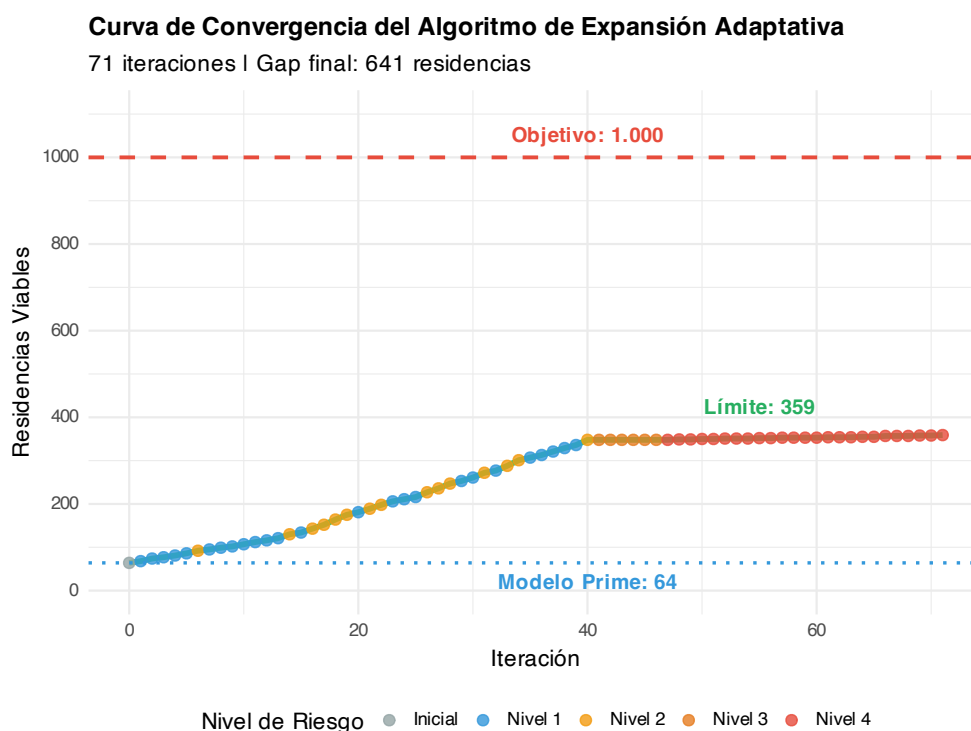


Figura 9: Curva de convergencia del algoritmo de expansión. Los colores indican el nivel de riesgo del parámetro relajado en cada iteración. La línea punteada azul representa el modelo Prime (64 residencias) y la línea discontinua roja el objetivo (1.000).

7.3.1. Parámetros Finales

Tras agotar todas las posibilidades de relajación, los parámetros finales fueron:

- **Percentil de Score:** 60 (antes 85)
- **Market Share (ϕ):** 6,0 % (antes 3,0 %)
- **Penalización Renta:** 0,70 (antes 0,40)
- **Camas Mínimas:** 60 (antes 85)

7.4. Comparativa: Modelo Prime vs Modelo Expandido

La Tabla 7 presenta la comparación detallada entre ambos modelos.

Cuadro 7: Comparativa entre Modelo Prime y Modelo Expandido

Métrica	Prime	Expandido	Δ
Residencias Potenciales	64	359	+295
Clusters Viables	21	91	+70
Camas Totales	6.440	35.991	+29.551
Renta Media Cluster (€)	42.517	40.410	-2.107
Score Demográfico Medio	0,594	0,470	-0,124
Tamaño Medio Residencia (camas)	306,7	395,5	+88,8

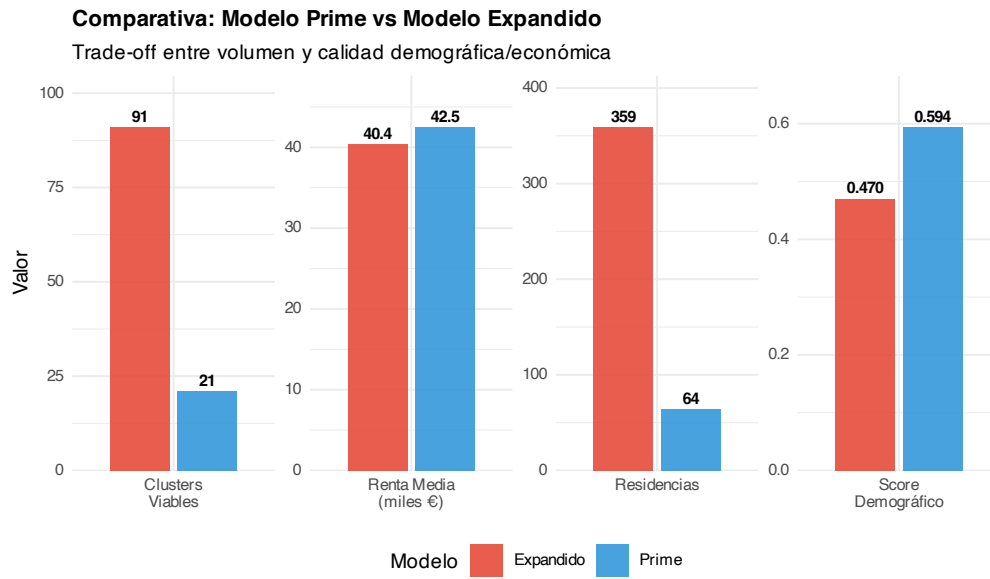


Figura 10: Comparativa visual entre el Modelo Prime (azul) y el Modelo Expandido (rojo). Se observa el trade-off entre volumen de residencias y calidad demográfica/económica.

7.5. Distribución Geográfica de Clusters Expandidos

La Figura 11 muestra la distribución espacial de los 91 clusters viables tras la expansión. El tamaño de cada punto representa la capacidad en camas, mientras que el color indica la renta media.



Figura 11: Distribución geográfica de los 91 clusters viables tras la expansión adaptativa. El tamaño representa la capacidad en camas y el color la renta media del cluster.

7.6. Interpretación del Gap Restante

A pesar de llevar todos los parámetros a sus límites de relajación, el modelo expandido alcanza únicamente **359 residencias**, dejando un **gap de 641 unidades** respecto al objetivo de 1.000.

7.7. Análisis de Rendimientos Decrecientes

Un análisis detallado de las 71 iteraciones revela un patrón de **rendimientos fuertemente decrecientes** que invalida la necesidad de llevar todos los parámetros a sus límites.

Cuadro 8: Eficiencia Marginal por Fase de Iteración

Fase	Iteraciones	Residencias	Δ	Ganancia/iter.	Parámetros Afectados
Fase 1	1–40	64 → 348	+284	7,1	Percentil, Market Share
Fase 2	41–71	348 → 359	+11	0,35	Penal. Renta, Camas Mín.

7.7.1. Punto de Inflexión: Iteración 40

En la iteración 40, el algoritmo alcanza **348 residencias** con los siguientes parámetros:

- Percentil de Score: 60 (límite)

- Market Share: 6,0 % (límite)
- Penalización Renta: **0,40** (sin modificar)
- Camas Mínimas: **85** (sin modificar)

A partir de este punto, las **31 iteraciones restantes** solo consiguen **+11 residencias adicionales** (+3,2 %), pero a costa de:

- Incrementar la tolerancia a rentas bajas de 0,40 a 0,70 (+75 %)
- Reducir el umbral de viabilidad de 85 a 60 camas (-29 %)

Esto representa una **eficiencia marginal 20 veces inferior** a la fase inicial (0,35 vs 7,1 residencias por iteración).

Recomendación Operativa

El punto óptimo de expansión se encuentra en la iteración 40, con 348 residencias potenciales. Continuar más allá de este punto ofrece ganancias marginales insignificantes (+11 residencias) a cambio de aceptar ubicaciones con menor solvencia económica y menor escala operativa. El modelo en la iteración 40 preserva la calidad del proyecto al mantener los umbrales de penalización por renta (0,40) y camas mínimas (85) en sus valores originales.

7.7.2. Conclusión sobre la Capacidad Real del Mercado

Conclusión Crítica

El análisis de frontera demuestra que **no existe capacidad demográfica suficiente** en el territorio español para soportar 1.000 nuevas residencias mediante construcción orgánica, incluso bajo supuestos agresivos de penetración de mercado (6 %) y umbrales reducidos de viabilidad (60 camas).

7.8. Implicaciones Estratégicas

1. **Límite estructural del mercado:** El gap de 641 residencias representa un límite demográfico-geográfico que no puede superarse únicamente con construcción de nuevas instalaciones.
2. **Deterioro de calidad:** Pasar del modelo Prime (64 residencias) al Expandido (359) implica aceptar una reducción del 21 % en el score demográfico medio y del 5 % en la renta media.
3. **Estrategia M&A imperativa:** Para cubrir el gap restante, es **ineludible** una estrategia de fusiones y adquisiciones de operadores existentes.
4. **Zona óptima:** El análisis sugiere que el punto de equilibrio entre volumen y calidad se encuentra entre 100-150 residencias, donde los parámetros aún mantienen niveles razonables de calidad.

8. Validación de Competencia: Análisis de Oferta Existente

8.1. Motivación del Análisis

El modelo L-SOMA identifica ubicaciones con **demanda potencial**, pero no considera la **oferta existente** de competidores. Una zona con alta demanda pero saturada de operadores no constituye una oportunidad real de negocio.

Para completar el análisis estratégico, se desarrolló un módulo de validación que cruza los clusters identificados con la oferta real de residencias obtenida de **Google Places API**.

8.2. Metodología

8.2.1. Parámetros del Análisis

Cuadro 9: Configuración del Análisis de Competencia

Parámetro	Valor	Justificación
Radio de búsqueda	1.500 m	Coherente con DBSCAN ($\varepsilon=1,5\text{km}$)
Camas por competidor	75	Media nacional (IMSERO 2023: 74,8 plazas/centro)
API utilizada	Text Search v1	Google Places API (New)
Query de búsqueda	"Residencia de ancianos OR Geriátrico"	Términos más frecuentes del sector

8.2.2. Cálculo del Índice de Saturación

Para cada cluster, se calcula el **Índice de Saturación** (I_{sat}):

$$I_{sat} = \frac{\text{Oferta Estimada (camas)}}{\text{Demanda Potencial (camas)}} = \frac{N_{competidores} \times 75}{\text{Camas_Potenciales}} \quad (18)$$

8.2.3. Clasificación Estratégica

Los clusters se clasifican según su I_{sat} :

- **Blue Ocean** ($I_{sat} < 0,20$): Mercado con poca o nula competencia. *Oportunidad premium.*
- **Batalla** ($0,20 \leq I_{sat} \leq 1,0$): Competencia moderada. Viable con diferenciación.
- **Saturado** ($I_{sat} > 1,0$): La oferta supera la demanda. *A descartar.*

8.3. Resultados

De los 91 clusters analizados:

Cuadro 10: Distribución de Clusters por Nivel de Saturación

Clasificación	Clusters	%	Interpretación
Blue Ocean	3	3,3 %	Oportunidades de oro
Batalla	13	14,3 %	Viabiles con estrategia
Saturado	75	82,4 %	Descartados
Total	91	100 %	—

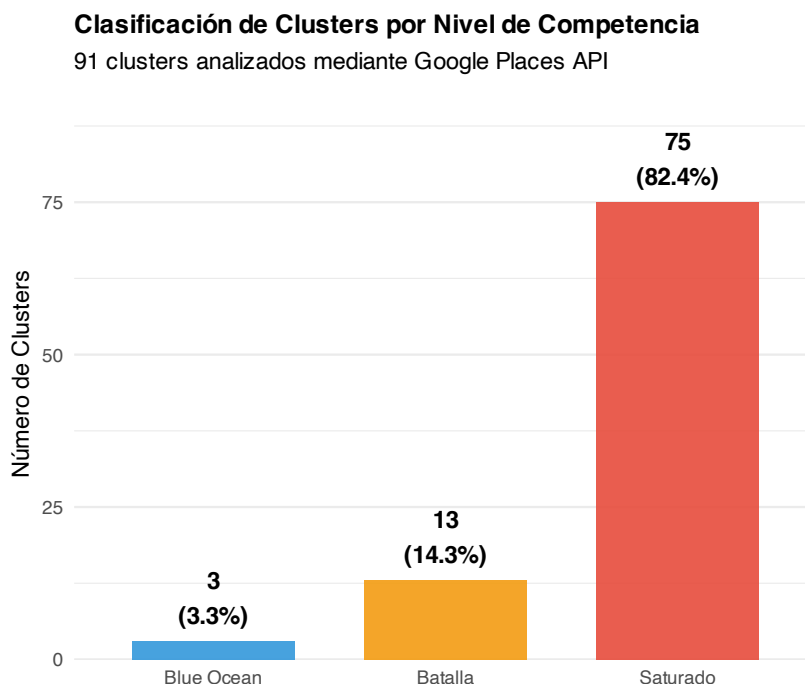


Figura 12: Distribución de los 91 clusters según nivel de saturación de mercado. Solo 3 clusters (3,3 %) califican como Blue Oceans.

8.4. Los 3 Blue Oceans Identificados

Solo 3 clusters presentan un índice de saturación inferior a 0,20, constituyendo las **oportunidades premium** del análisis:

Cuadro 11: Blue Oceans: Clusters con Mínima Competencia

Rank	Ubicación	Camas	Comp.	I_{sat}	Rival cercano
1	Madrid Centro	7.274	7	0,077	Hermanitas Desamparados (230m)
2	Bilbao	2.114	4	0,151	Anai Cantero (367m)
3	Málaga Costa	419	1	0,191	Hermanitas Pobres (1.041m)

8.4.1. Análisis de los Blue Oceans

1. **Madrid Centro (Cluster 4):** Con 1.577 secciones censales y una demanda potencial de 7.274 camas, presenta solo 7 competidores identificados. El I_{sat} de 0,077 indica que la oferta actual cubre apenas el 7,7% de la demanda potencial. Sin embargo, este cluster presenta una **alta barrera de entrada inmobiliaria**: el elevado coste del suelo urbano hace inviable el modelo greenfield tradicional. La oportunidad reside en la *reconversión* de edificios de otros usos (oficinas, hoteles) o en residencias de *alto standing* que justifiquen el repercutido del suelo.
2. **Bilbao (Cluster 12):** Área metropolitana con 476 secciones y 2.114 camas potenciales. Los 4 competidores detectados dejan un gap de más de 1.800 camas, indicando una **necesidad social no cubierta**. Al igual que Madrid, requiere un operador institucional con capacidad de CAPEX elevado para desbloquear suelo urbano o proyectos de reconversión.
3. **Málaga Costa (Cluster 27):** Zona con 110 secciones, renta media de 40.418€ y un único competidor a más de 1km del centroide. Menor barrera de entrada y potencial para modelo greenfield tradicional. Representa una **oportunidad de nicho** para operadores medianos.

8.5. Limitaciones del Análisis de Competencia

1. **Cobertura de Google Places:** La API puede no incluir todos los operadores, especialmente centros pequeños o recién abiertos.
2. **Estimación de camas:** Se asume 75 camas/competidor. La realidad varía entre 30 y 200 camas por centro.
3. **No se considera ocupación:** Un competidor puede estar al 50% de ocupación, dejando demanda sin cubrir.
4. **Snapshot temporal:** Los datos reflejan el estado a enero 2026. El mercado es dinámico.

9. Discusión

9.1. Síntesis de Hallazgos Cuantitativos

El análisis exhaustivo del territorio español mediante el algoritmo L-SOMA ha producido los siguientes hallazgos verificados:

Cuadro 12: Métricas Principales del Análisis L-SOMA

Métrica	Valor Verificado
Total secciones censales analizadas	32.910
Secciones en Top 15 % (candidatas)	4.937
Clusters detectados (DBSCAN $\varepsilon=1,5\text{km}$)	135
Clusters viables (Modelo Prime, ≥ 85 camas)	21
Residencias potenciales (Modelo Prime, 3 %)	64
Residencias máximo teórico (Modelo Expandido, 6 %)	359
Target promedio real (mujeres 80+ /sección)	70,1
<i>Validación de Competencia (Google Places):</i>	
Clusters Blue Ocean ($I_{sat} < 0, 20$)	3 (3,3 %)
Clusters Batalla ($0, 20 \leq I_{sat} \leq 1, 0$)	13 (14,3 %)
Clusters Saturados ($I_{sat} > 1, 0$)	75 (82,4 %)

9.2. Concentración Geográfica de la Demanda

Los datos revelan una **fuerte concentración metropolitana** de la demanda viable. Los 5 clusters con mayor potencia acumulan el 56 % de la capacidad total:

1. **Madrid (Cluster 3):** 722 secciones, Potencia 421,7, Renta media 47.466€
2. **Barcelona (Cluster 4):** 323 secciones, Potencia 185,5, Renta media 45.313€
3. **Bilbao (Cluster 11):** 209 secciones, Potencia 124,3, Renta media 43.073€
4. **Zaragoza (Cluster 2):** 141 secciones, Potencia 83,1, Renta media 42.624€
5. **Valladolid (Cluster 12):** 98 secciones, Potencia 59,5, Renta media 40.330€

Nota: Los IDs de cluster corresponden al ranking demográfico (`ranking_fase7_clusters.csv`). El análisis de competencia posterior (Sección 8) opera sobre el dataset expandido, donde los IDs difieren.

El restante 44 % se distribuye entre 130 clusters de menor escala, muchos de los cuales no alcanzan el umbral de viabilidad operativa.

9.3. Análisis del Gap Estratégico

El objetivo inicial de 1.000 residencias presenta un **gap estructural insalvable** mediante construcción orgánica:

- **Modelo Prime (parámetros estrictos):** 64 residencias viables
- **Modelo Expandido (parámetros relajados):** 359 residencias máximo
- **Gap irreducible:** 641 residencias (64 % del objetivo)

El algoritmo de expansión adaptativa demostró que, incluso llevando todos los parámetros a sus límites (percentil 60, market share 6 %, penalización 0,70, camas mínimas 60), el mercado español no soporta más de 359 ubicaciones potenciales.

9.4. Rendimientos Decrecientes y Punto Óptimo

El análisis de las 72 iteraciones del algoritmo de expansión (0–71) reveló un patrón crítico:

- **Iteraciones 1-40:** +284 residencias (7,1 por iteración)
- **Iteraciones 41-71:** +11 residencias (0,35 por iteración)

La eficiencia marginal cae un 95 % tras la iteración 40, donde se alcanzan 348 residencias sin degradar los parámetros de calidad económica (penalización renta y camas mínimas). Este constituye el **punto óptimo operativo**.

9.5. Saturación del Mercado: Análisis de Competencia

La validación mediante Google Places API reveló un dato crítico: **el 82,4 % de los clusters viables ya están saturados** de competidores.

- **Blue Oceans (3 clusters):** Madrid Centro (7.274 camas, $I_{sat}=0,077$), Bilbao (2.114 camas, $I_{sat}=0,151$) y Málaga Costa (419 camas, $I_{sat}=0,191$). Son las únicas oportunidades reales de entrada sin competencia directa.
- **Batallas (13 clusters):** Mercados viables pero con competencia moderada. Requieren diferenciación o ventaja competitiva.
- **Saturados (75 clusters):** La oferta existente supera la demanda potencial. Desaconsejados para nueva construcción.

Este hallazgo reduce drásticamente el universo de oportunidades: de las 64-348 residencias teóricamente viables, solo las ubicadas en los **3 Blue Oceans** garantizan entrada sin competencia directa.

9.6. Limitaciones del Estudio

1. **Horizonte temporal estático:** Los datos corresponden al censo 2021-2025. No se modelan proyecciones demográficas futuras. Una sección con muchos habitantes de 60-65 años hoy será un *hotspot* en 15 años, pero el modelo actual la descarta. Para inversiones inmobiliarias con amortización a 20-30 años, esta es una limitación relevante.
2. **Cuota de mercado estimada:** Los escenarios de 1,5 % a 6 % de market share son hipotéticos; no existen datos públicos de tasas de institucionalización por segmento premium.
3. **Incertidumbre en la competencia (Google Places):** La estimación del índice de saturación I_{sat} depende de dos asunciones frágiles: (1) que la API de Google captura el 100 % de la oferta relevante (falso en zonas rurales o centros antiguos), y (2) que cada competidor tiene 75 camas (media nacional). La varianza real oscila entre 20 y 200 camas por centro, introduciendo un margen de error significativo en la clasificación Blue Ocean vs. Saturado.

4. **Proxy de presión de cuidados (Π_i):** El ratio asume que todas las mujeres de 45-64 años en la sección son cuidadoras potenciales, ignorando la movilidad laboral (hijas que residen en otro código postal). Es un proxy de proximidad, no de disponibilidad real para el cuidado.
5. **Target demográfico simplificado:** El vector Q considera mujeres 80+ de hogares >30k€ como proxy de demanda premium, excluyendo otros segmentos potencialmente interesados.

10. Conclusiones

El desarrollo y ejecución del algoritmo L-SOMA sobre las 32.910 secciones censales españolas permite extraer las siguientes conclusiones fundamentadas en datos:

10.1. Validación Metodológica

1. **Divergencia Jensen-Shannon efectiva:** La métrica JSD permitió ordenar las secciones por “resonancia” con el perfil objetivo, reduciendo el universo de análisis de 32.910 a 4.937 candidatas (Top 15 %).
2. **DBSCAN geodésico funcional:** El clustering espacial con $\varepsilon=1,5\text{km}$ identificó 135 agrupaciones naturales de demanda, de las cuales 21 superan el umbral de viabilidad en el modelo Prime.
3. **Auditoría de supuestos exitosa:** La corrección del target promedio de 110 a 70,1 mujeres/sección (verificado desde Matriz P) demostró la importancia de validar constantes con datos reales.

10.2. Hallazgos Estratégicos

1. **Capacidad máxima verificada:** España soporta entre **64 y 348 residencias** mediante construcción orgánica, dependiendo del nivel de exigencia en los parámetros de calidad.
2. **Concentración metropolitana:** Los 5 primeros clusters (Madrid, Barcelona, Bilbao, Zaragoza, Valladolid) concentran el 56 % de la potencia total identificada.
3. **Objetivo de 1.000 residencias inalcanzable orgánicamente:** El gap de 641 unidades requiere obligatoriamente una estrategia de fusiones y adquisiciones (M&A).
4. **Punto óptimo en iteración 40:** Con 348 residencias se maximiza el volumen preservando la calidad del modelo (eficiencia marginal cae 95 % más allá de este punto).
5. **Mercado altamente saturado:** El 82,4 % de los clusters viables presenta saturación de oferta ($I_{sat} > 1$). Solo 3 clusters (Madrid Centro, Bilbao, Málaga Costa) califican como Blue Oceans.

6. **Madrid Centro: oportunidad excepcional:** Con 7.274 camas potenciales y solo 7 competidores identificados ($I_{sat}=0,077$), representa la mayor oportunidad teórica, condicionada a la alta barrera de entrada inmobiliaria (disponibilidad de activos y coste del suelo)..

10.3. Recomendación Estratégica Final

Basándose exclusivamente en los datos analizados, la estrategia óptima combina:

- **Construcción orgánica selectiva:** 64-100 residencias en los clusters de máxima potencia (Madrid, Barcelona, Bilbao), aplicando el modelo Prime para garantizar calidad demográfica.
- **Expansión moderada:** Hasta 250 residencias adicionales relajando el percentil a 60 y el market share a 6 %, pero preservando penalización de renta (0,40) y camas mínimas (85).
- **M&A para el gap restante:** Las 641 residencias faltantes solo pueden cubrirse mediante adquisición de operadores existentes fuera del alcance de este modelo.

A. Anexo A: Datos de Entrada

El análisis L-SOMA utilizó las siguientes fuentes de datos oficiales:

Cuadro 13: Fuentes de Datos Utilizadas

Dato	Fuente	Año
Población por edad/sexo	INE - Censo de Población	2021-2024
Renta media por hogar	INE - Atlas de Renta	2023
Geometría secciones censales	IGN - SECC_CE_20240101	2024
Coordenadas centroides	Cálculo propio desde shapefile	2024

Vector Q (Target demográfico): Mujeres mayores de 80 años en hogares con renta >30.000€/año. Representan el 6,2 % de la población según Matriz P.

Poblaciones procesadas:

- Secciones censales totales: 32.910
- Secciones con coordenadas válidas: 32.910
- Secciones en Top 15 % por Score: 4.937
- Secciones asignadas a clusters: 2.624
- Secciones clasificadas como ruido: 844

B. Anexo B: Pipeline de Procesamiento

El proyecto se estructura en scripts Python ejecutados secuencialmente:

Cuadro 14: Pipeline de Scripts del Algoritmo L-SOMA

Script	Función
07_definicion_target_v3.py	Genera el vector objetivo $ Q\rangle$ (42 dimensiones)
08_generador_matriz_P_v2.py	Construye matriz demográfica \mathbf{P} (32.910×42)
09_auditoria_matriz.py	Valida estructura, masa y normalización
10_calculo_resonancia.py	Calcula JSD para 32.910 secciones
11_ingesta_renta_v2.py	Integra datos económicos (renta hogar)
12_feature_engineering_hijas.py	Calcula ratios hijas/abuelas y presión cuidados
13_preparacion_geoespacial.py	Añade coordenadas LAT/LON
14_clustering_demanda.py	Ejecuta DBSCAN ($\varepsilon=1,5\text{km}$, $\text{min}=3$)
15_calculo_masa_critica.py	Evalúa viabilidad por cluster
19_informe_final_corregido.py	Genera resultados auditados
expansion_1000_residencias.py	Algoritmo de expansión adaptativa (71 iteraciones)
validacion_competencia_V4.py	Análisis de competencia vía Google Places API

C. Anexo C: Parámetros del Modelo

C.1. Parámetros Fijos (No Modificados)

- Radio DBSCAN (ε): 1,5 km (haversine)
- Mínimo secciones por cluster: 3
- Columnas target (Vector Q): M_80-84, M_85-89, M_90-94, M_95-99, M_100+
- Umbral renta solvencia: 30.000 €/año

C.2. Parámetros Variables (Escenarios)

Cuadro 15: Espacio de Parámetros Explorado

Parámetro	Prime	Expandido	Rango
Percentil Score	85	60	60–85
Market Share (ϕ)	3,0 %	6,0 %	1,5 %–6,0 %
Penalización Renta	0,40	0,70	0,40–0,70
Camas Mínimas	85	60	60–85

C.3. Código: Divergencia Jensen-Shannon

Listing 1: Cálculo de resonancia demográfica

```
from scipy.spatial.distance import jensenshannon
import numpy as np

# vector_Q: probabilidades target (42 dims)
# matriz_P_np: PDFs por seccion (32910 x 42)

distancias = np.apply_along_axis(
    lambda row: jensenshannon(row, vector_Q),
    axis=1,
    arr=matriz_P_np
)
resonancias = 1.0 - distancias
```

C.4. Código: Clustering DBSCAN Geodésico

Listing 2: Configuración de DBSCAN

```
from sklearn.cluster import DBSCAN

RADIO_KM = 1.5
MIN_SECCIONES = 3
```



```

coords_rad = np.radians(df[['LATITUD', 'LONGITUD']])
epsilon = RADIO_KM / 6371.0 # radianes

db = DBSCAN(
    eps=epsilon,
    min_samples=MIN_SECCIONES,
    metric='haversine',
    algorithm='ball_tree'
)
df['Cluster_ID'] = db.fit_predict(coords_rad)

```

D. Anexo D: Validación de Competencia

D.1. Configuración de Google Places API

Cuadro 16: Parámetros de la API de Competencia

Parámetro	Valor
Endpoint	places.googleapis.com/v1/places:searchText
Radio de búsqueda	1.500 metros
Máximo resultados	20 por cluster
Query	“Residencia de ancianos OR Geriátrico”
Camas estimadas/competidor	75 (media IMSERSO)

D.2. Filtros de Calidad Aplicados

Para evitar falsos positivos, se implementaron dos filtros:

1. Blacklist (descarte por nombre):

- farmacia, ortopedia, gimnasio, club, asociación, parking, ayuntamiento

2. Whitelist (validación positiva):

- **Por tipo Google:** nursing_home, assisted_living_complex, health
- **Por nombre:** residencia, geriátrico, mayores, tercera edad

D.3. Archivos Generados

- clusters_359_validado_FINAL.csv: 91 clusters con métricas de competencia
- Columnas añadidas: Num_Competidores, Rivales_Cercanos, Distancia_Media, Indice_Saturacion, Tipo_Oceano

Referencias

- [1] Instituto Nacional de Estadística. *Censo de Población y Viviendas 2021*. INE, Madrid, 2023.
- [2] Instituto Nacional de Estadística. *Atlas de Distribución de Renta de los Hogares*. INE, Madrid, 2023.
- [3] Lin, J. (1991). Divergence measures based on the Shannon entropy. *IEEE Transactions on Information Theory*, 37(1), 145-151.
- [4] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters. *Proceedings of KDD-96*, 226-231.
- [5] CSIC - Enred. *¿Quién cuida a quién? La disponibilidad de cuidadores informales*. <http://envejecimiento.csic.es/documentos/documentos/enred-info20-quiencuida.pdf>
- [6] Envejecimiento en Red. *Informe 2022: El perfil de las personas mayores en España*. https://envejecimientoenred.github.io/informe_perfil_mayores_2022_social_4_4/
- [7] Murcia Salud. *El aumento de la esperanza de vida retrasa la edad de ingreso en residencias*. <https://www.murciasalud.es/-/el-aumento-de-la-esperanza-de-vida-retrasa-la-edad-de-ingreso-de-los-mayores-en->
- [8] INE. *Encuesta Continua de Hogares - Últimos datos*. https://www.ine.es/dyngs/INEbase/es/operacion.htm?c=Estadistica_C&cid=1254736176952
- [9] Matas Pino, A. *Las consecuencias del síndrome de burnout en cuidadores familiares*. Universidad Pontificia Comillas, TFG, 2024.
- [10] Sanitas. *Síndrome del cuidador quemado: qué es, causas y soluciones*. <https://www.sanitas.es/biblioteca-de-salud/tercera-edad/habitos-vida-saludable/sindrome-del-cuidador-quemado>
- [11] CSIC. *Protocolo de ingreso en centros residenciales*. <http://envejecimiento.csic.es/documentos/documentos/chs-jornadas2003-01.pdf>
- [12] Ballesol. *Los síndromes geriátricos y la salud*. <https://ballesol.es/sindromes-geriatricos/>
- [13] Valdeluz. *¿Cómo elegir residencia de mayores? 10 factores a considerar*. <https://www.valdeluz.com/noticias/como-elegir-residencia-mayores/>
- [14] Universidad de Barcelona. *Las barreras arquitectónicas y las personas mayores*. *Scripta Nova*, Vol. X, núm. 218 (104).
- [15] Comunidad de Madrid. *Edificios destinados a viviendas con accesibilidad y sin ascensor*. Portal de datos abiertos, 2024.
- [16] Inforesidencias. *Informe de precios 2025 de las residencias de mayores de España*. <https://www.inforesidencias.com/contenidos/noticias/nacional/informe-de-precios-del-2025>

- [17] Grupo Emera. *Requisitos para ingresar en una residencia de mayores.* <https://emera-group.es/noticias/requisitos-para-ingresar-en-una-residencia-de-mayores/>
- [18] Residenciatura. *Cuidar de un familiar mayor: ¿ha llegado el momento de pensar en una residencia?* <https://www.residenciatura.org/es/article/cuidar-familiar-mayor-ha-llegado-momento-pensar-residencia>
- [19] Sanitas. *Cuidado de seres queridos: ¿cómo gestionar el sentimiento de culpa?* <https://www.sanitas.es/biblioteca-de-salud/tercera-edad/hogar/cuidar-con-culpa-una-mala-combinacion>
- [20] Minut. *Coliving senior España: guía completa del modelo para mayores.* <https://www.minut.com/es/blog/coliving-senior-espana>
- [21] Culmia. *Senior Living: Nuevo modelo de viviendas para la tercera edad.* <https://www.culmia.com/es/blog/senior-living-modelo-de-viviendas-tercera-edad/>
- [22] Sanitas. *10 principales ventajas de vivir en una residencia de mayores.* <https://www.sanitas.es/biblioteca-de-salud/tercera-edad/hogar/diez-ventajas-de-vivir-en-una-residencia-de-ancianos>
- [23] Ballesol. *Información y Orientación para familias.* <https://ballesol.es/informacion-y-orientacion/>
- [24] Los Nogales. *La cercanía, un factor clave en la elección del mejor entorno.* <https://los-nogales.es/cercania-de-la-residencia-de-ancianos-un-factor-clave/>
- [25] Resimas. *Las 10 ventajas de escoger una residencia de mayores cerca de casa.* <https://resimas.com/las-10-ventajas-de-escoger-una-residencia-de-mayores-cerca-de-casa/>