

Análisis de Clustering

SEDRONAR

LÍNEA 141

Primer Semestre 2025

Realizado por Natalia G. Cuellas



ÍNDICE

01

Descripción

02

Contexto

03

Valor del análisis

04

Necesidad a resolver

05

Modelo de datos

06

Decisiones de diseño

07

Resultado

08

Conclusiones



DESCRIPCIÓN

La **Secretaría de Políticas Integrales sobre Drogas de la Nación Argentina** (SEDRONAR) es el organismo nacional que coordina políticas públicas de prevención, asistencia y control de adicciones en Argentina.

Su programa **Línea 141** ofrece atención gratuita, confidencial y disponible las 24 horas, brindando escucha, orientación y derivación ante situaciones de consumo problemático.

Atiende tanto consultas directas (personas que buscan ayuda) como indirectas (familiares o allegados), siendo un servicio clave de contención y acompañamiento en todo el país.



CONTEXTO

- **Distribución de llamadas**

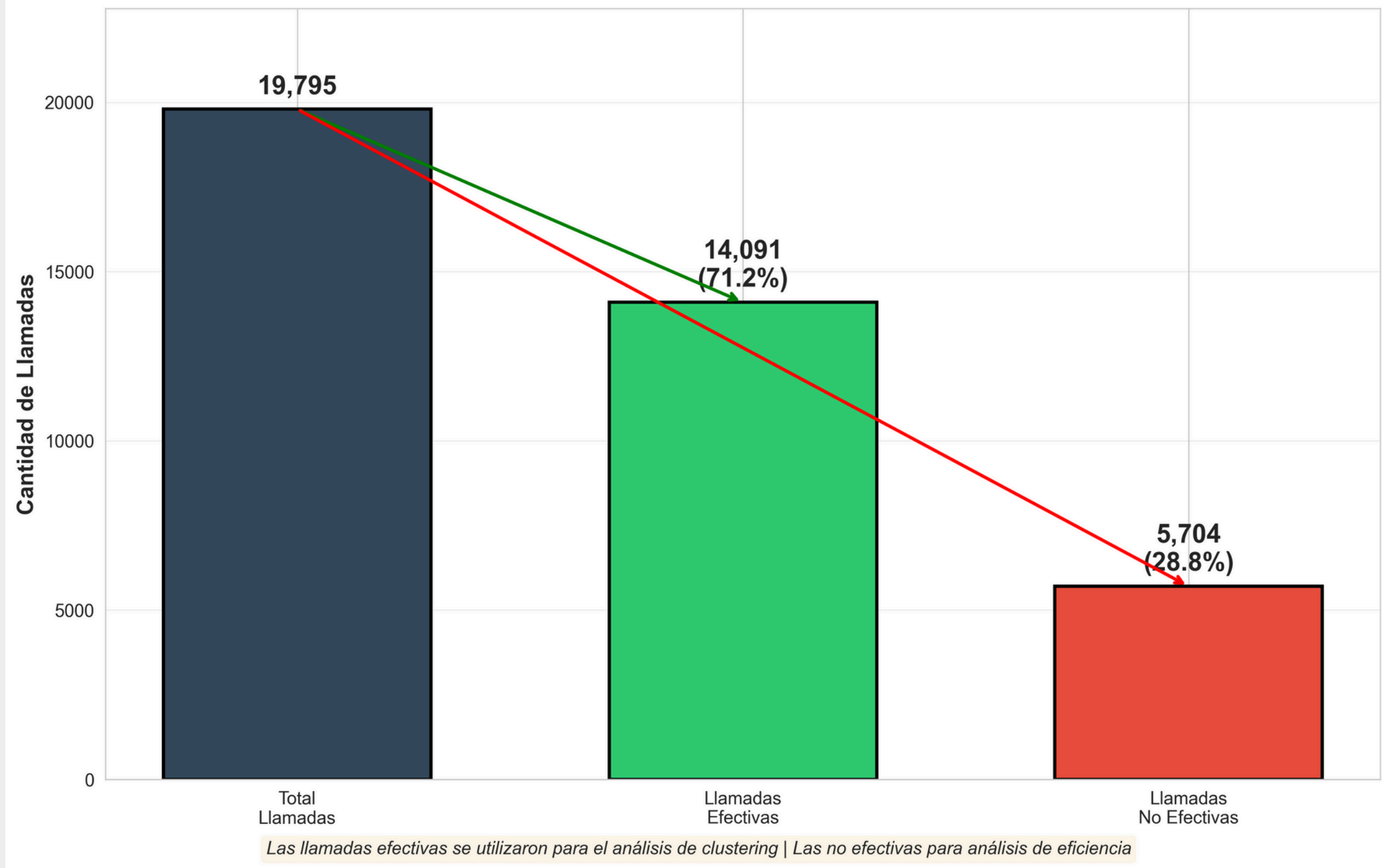
Durante el primer semestre de 2025, la Línea 141 recibió más de 14.000 llamadas en todo el país, con mayor concentración en la Provincia de Buenos Aires ($\approx 58\%$) y la Ciudad Autónoma de Buenos Aires ($\approx 12\%$). Estos datos evidencian el alcance nacional del servicio y la necesidad de analizar su uso para fortalecer las políticas públicas de prevención y asistencia.

- **Problemática social**

Las consultas reflejan una problemática social compleja, vinculada tanto a la demanda de asistencia directa por parte de personas en consumo, como a la búsqueda de orientación de familiares o allegados. Este patrón muestra la diversidad de situaciones abordadas por la Línea 141 y su relevancia como herramienta de contención y ayuda.



Segmentación del Dataset - Línea 141 (1er Semestre 2025)



El dataset contiene 19,795 llamadas recibidas en el primer semestre de 2025. Se segmentaron en 14,091 llamadas efectivas (71.2%) para el análisis de clustering y 5,704 llamadas no efectivas (28.8%) para análisis de eficiencia del servicio.

VALOR DEL ANÁLISIS

El estudio de estos datos permite a SEDRONAR identificar patrones de uso, perfiles de consultantes y optimizar la asignación de recursos humanos y comunicacionales. Desde la gestión pública, la Línea 141 se consolida como una fuente estratégica de información para diseñar acciones basadas en evidencia y mejorar la respuesta estatal ante el consumo problemático.

NECESIDAD

La Línea 141 recibe diariamente consultas de todo el país, pero la información no siempre se utiliza de forma analítica.

Existe la necesidad de identificar patrones y perfiles de llamadas para optimizar la atención, mejorar la asignación de recursos y diseñar estrategias preventivas más efectivas frente al consumo problemático de sustancias.



MODELO DE DATOS



El dataset de la Línea 141 recopila información estructurada de cada llamada recibida en distintos tipos de datos

- **Temporales**

Incluyen la fecha, día y hora de cada llamada. Permiten detectar momentos de mayor demanda, analizar patrones semanales y planificar la disponibilidad del servicio según los horarios críticos.

- **Demográficos**

Contienen datos de edad, género y provincia del consultante o consumidor. Ayudan a identificar perfiles poblacionales y zonas con mayor incidencia para orientar acciones preventivas.

- **Contextuales**

Abarcan el motivo, tipo y modalidad de la consulta. Reflejan si la llamada fue directa o indirecta y el nivel de urgencia, aportando información clave sobre las necesidades de atención.



Técnica Aplicada

Se utilizó **Clustering (agrupamiento no supervisado)**, ya que el dataset no posee variable objetivo. El propósito fue descubrir patrones y agrupar llamadas con características similares, identificando perfiles de consultantes y tipos de demanda para comprender mejor el servicio y optimizar la asignación de recursos.



Pipeline ETL

Extracción: Dataset público (Sedronar), con 19.795 registros del primer semestre 2025 (14.091 llamadas efectivas).

Transformación: Limpieza de caracteres, conversión de fechas, creación de variables derivadas, estandarización de categorías, codificación One-Hot e imputación de faltantes.

Carga: Integración final en DataFrames segmentados (efectivas/no efectivas) para modelado y análisis exploratorio.



Algoritmo seleccionado

Se aplicó K-Means, adecuado para datos mixtos preprocesados con One-Hot Encoding. El modelo busca minimizar la inercia intra-cluster y maximizar la separación entre grupos. El número de clústeres se definió con el método del codo y el índice silhouette, asegurando una segmentación coherente, reduciendo el ruido dimensional.



Stack Tecnológico

Lenguaje: Python 3.12

Librerías: pandas, numpy, scikit-learn, matplotlib, seaborn

Entorno: Jupyter Notebook en Visual Studio Code

Control de versiones: GitHub (tp-mineria-141)

Salidas: Gráficos de clúster exportados a PNG (outputs/),
tablas CSV con perfiles y métricas.

01

Se espera generar segmentos homogéneos de llamadas que revelen patrones ocultos tales como: consultas informativas sin caso específico; consumidores adultos crónicos con historial de recaídas, jóvenes con familias preocupadas en primeros contactos.

02

El valor es intentar interpretar perfiles de consultantes y necesidades diferenciadas; optimizar asignación de recursos por franja horaria y tipo de demanda; orientar políticas públicas basadas en evidencia e identificar grupos de riesgo para intervención prioritaria.

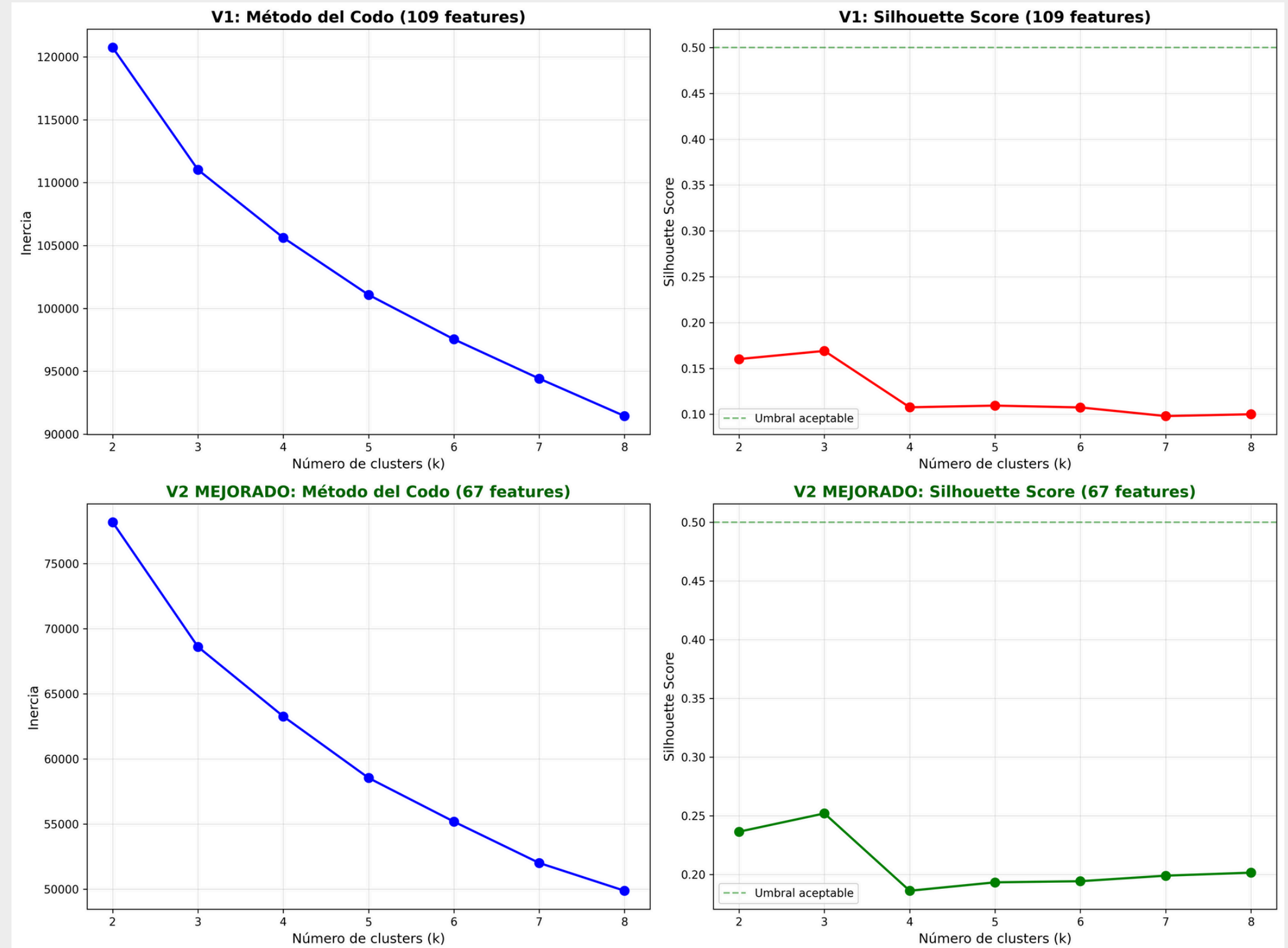
03

Se desea proveer una base reproducible para análisis longitudinales trimestrales/anuales. Input para modelos predictivos (recaídas, abandono de tratamiento), evaluación de eficiencia del servicio mediante llamadas no efectivas.

Se probaron dos configuraciones del modelo:

- V1: 109 features (silhouette: 0.17) - mucho ruido dimensional
- V2 (mejorado): 67 features (silhouette: 0.25) - mejora del 47%

El método del codo y el índice de Silhouette coincidieron en $k=3$ como número óptimo de clusters.






SEGMENTACIÓN

El modelo K-Means generó 3 grupos principales de llamadas, definidos por características demográficas, tipo de consulta, historial de tratamiento y patrones temporales.

Cada clúster representa un perfil diferenciado con necesidades específicas, lo que permite comprender quiénes llaman, por qué motivo, en qué momento y qué tipo de intervención requieren.



CLÚSTERS



- **Consultas informativas (41%)**

Agrupación de pedidos de información general sin referencia a un caso específico de consumo. Útiles para evaluar necesidad de material educativo y mejorar canales de atención automática.

- **Consumidores Adultos con historial (22.5%)**

Integrado por adultos de mediana edad (42 años promedio) con consumo prolongado (>10 años) e historial de tratamientos previos. Presentan recaídas y trastornos crónicos.

- **Jóvenes - Llamadas de Familiares (36.4%)**

Predominan jóvenes adultos (26 años promedio) por quienes consultan principalmente sus madres. Representan crisis recientes y primeros contactos con el sistema de atención.

CLÚSTERS



- **Consultas informativas (41%)**

Características principales:

- 100% tipo "Informe"
- Sin datos demográficos del consumidor
- Consultantes variados: consumidores, familiares, otros
- Horario: tarde principalmente

Acción sugerida: Crear guías automatizadas, FAQs, chatbots para derivación eficiente.

- **Consumidores Adultos con historial (22.5%)**

Características principales:

- Edad media: 42 años
- Tiempo de consumo: más de 10 años
- Ya tuvieron tratamientos previos
- Motivos: crisis, recaídas, trastornos crónicos

Acción sugerida: Programas de seguimiento prolongado, grupos de apoyo, prevención de recaídas.

- **Jóvenes - Llamadas de Familiares (36.4%)**

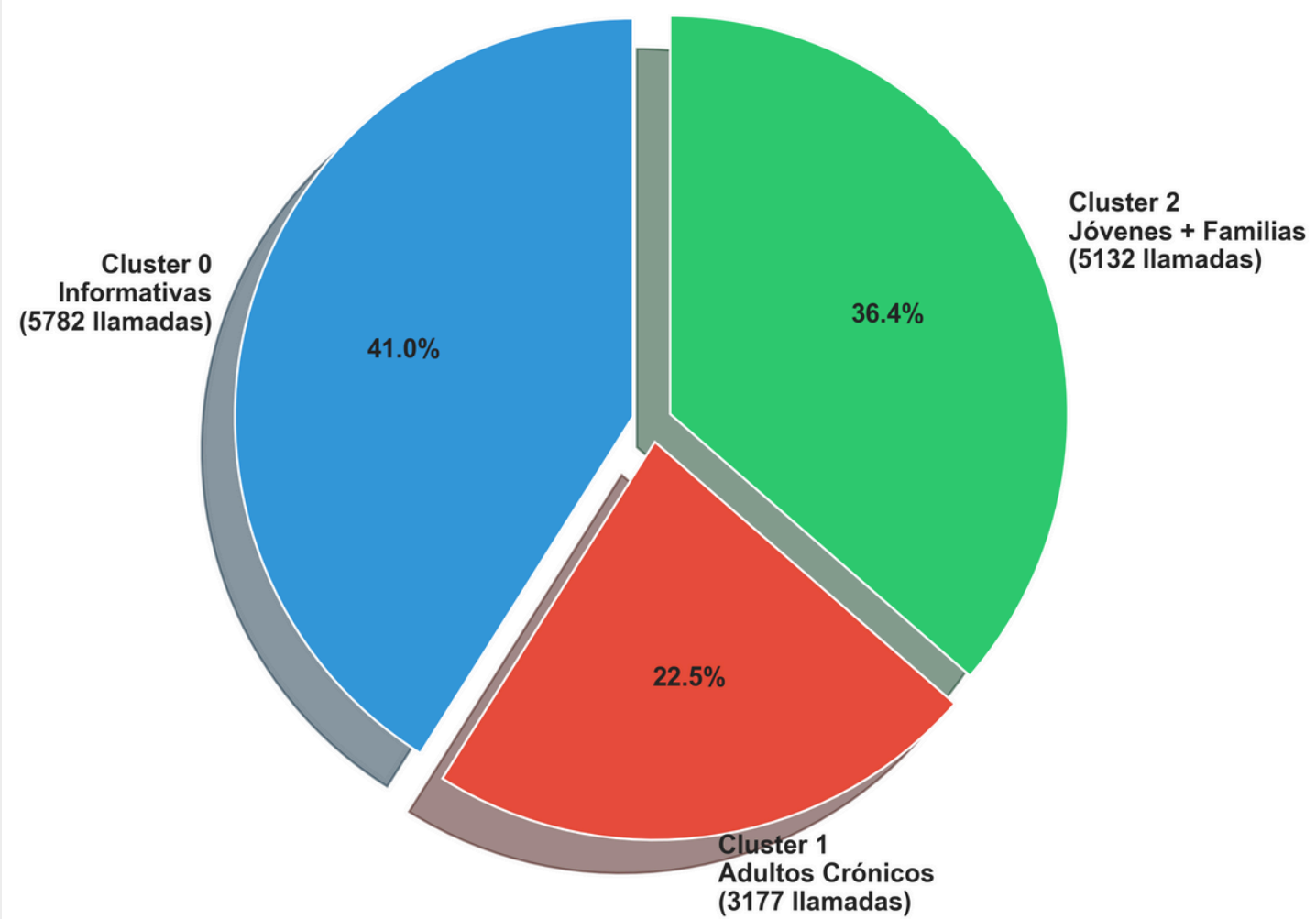
Características principales:

- Edad media: 26 años
- Consultante principal: Madre (patrón marcado)
- Consultas indirectas
- Tratamiento: nunca estuvieron o sin definir

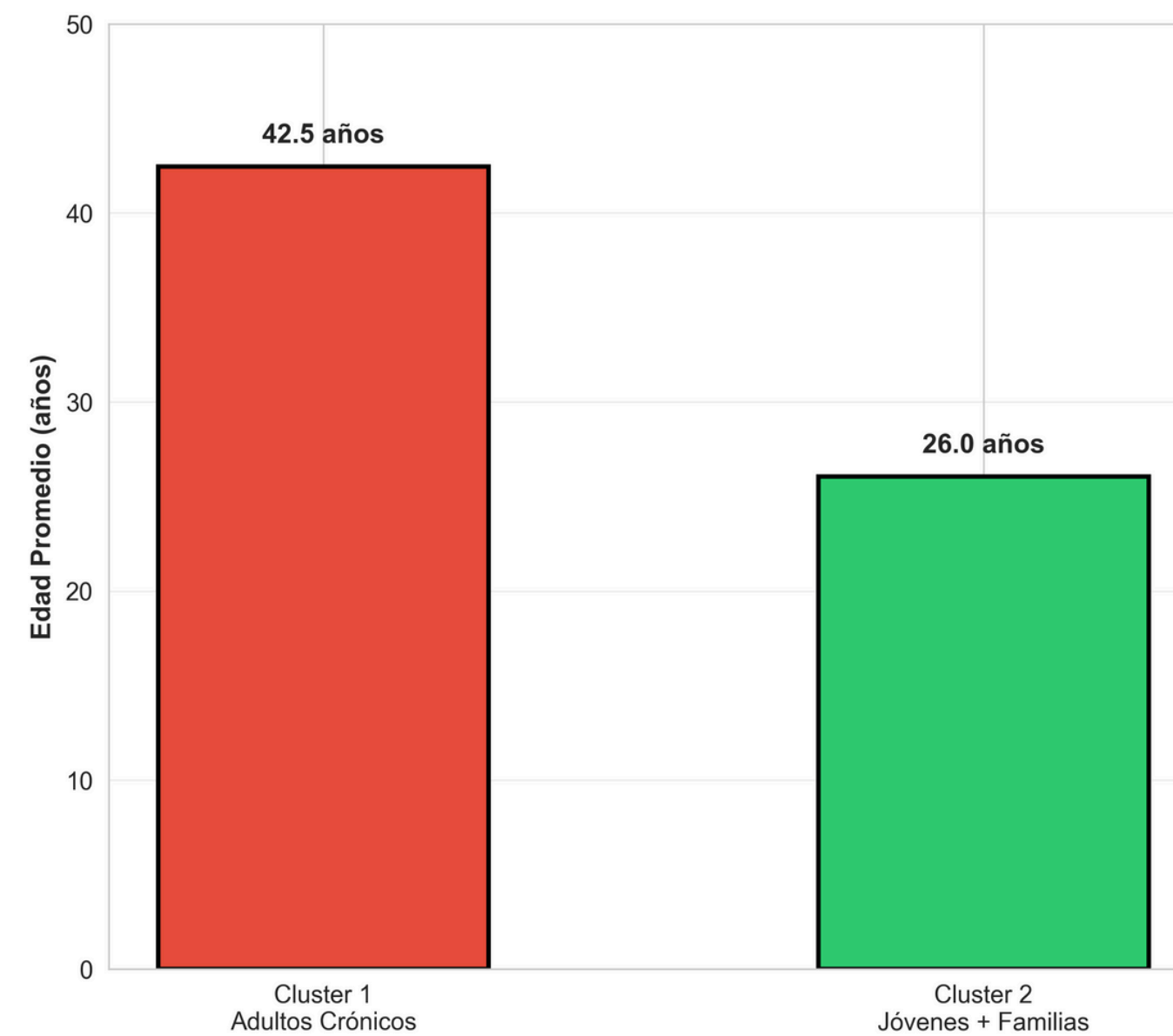
Acción sugerida: Orientación familiar, contención inmediata, derivación rápida a tratamiento.

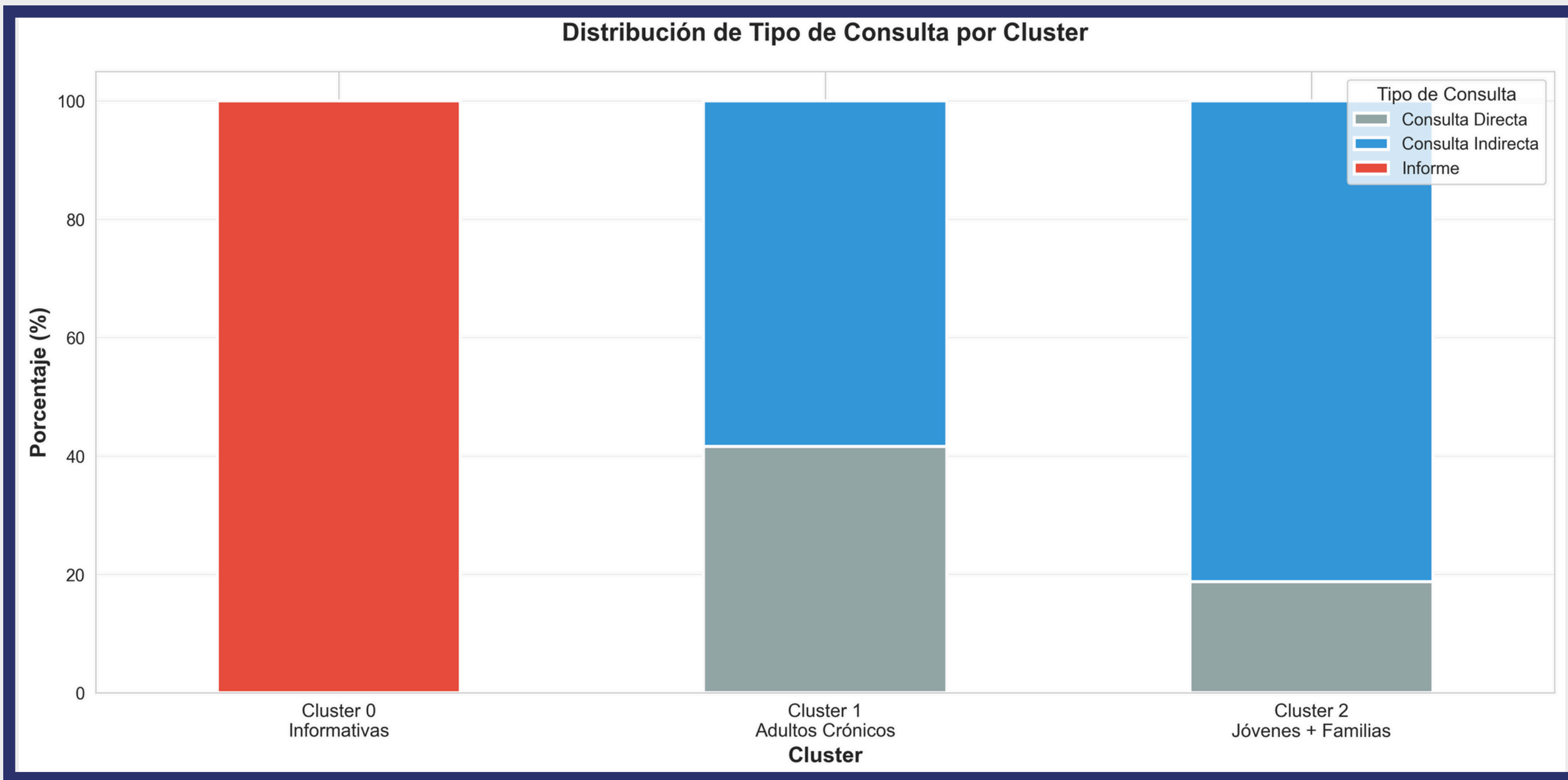


Distribución de Llamadas por Cluster



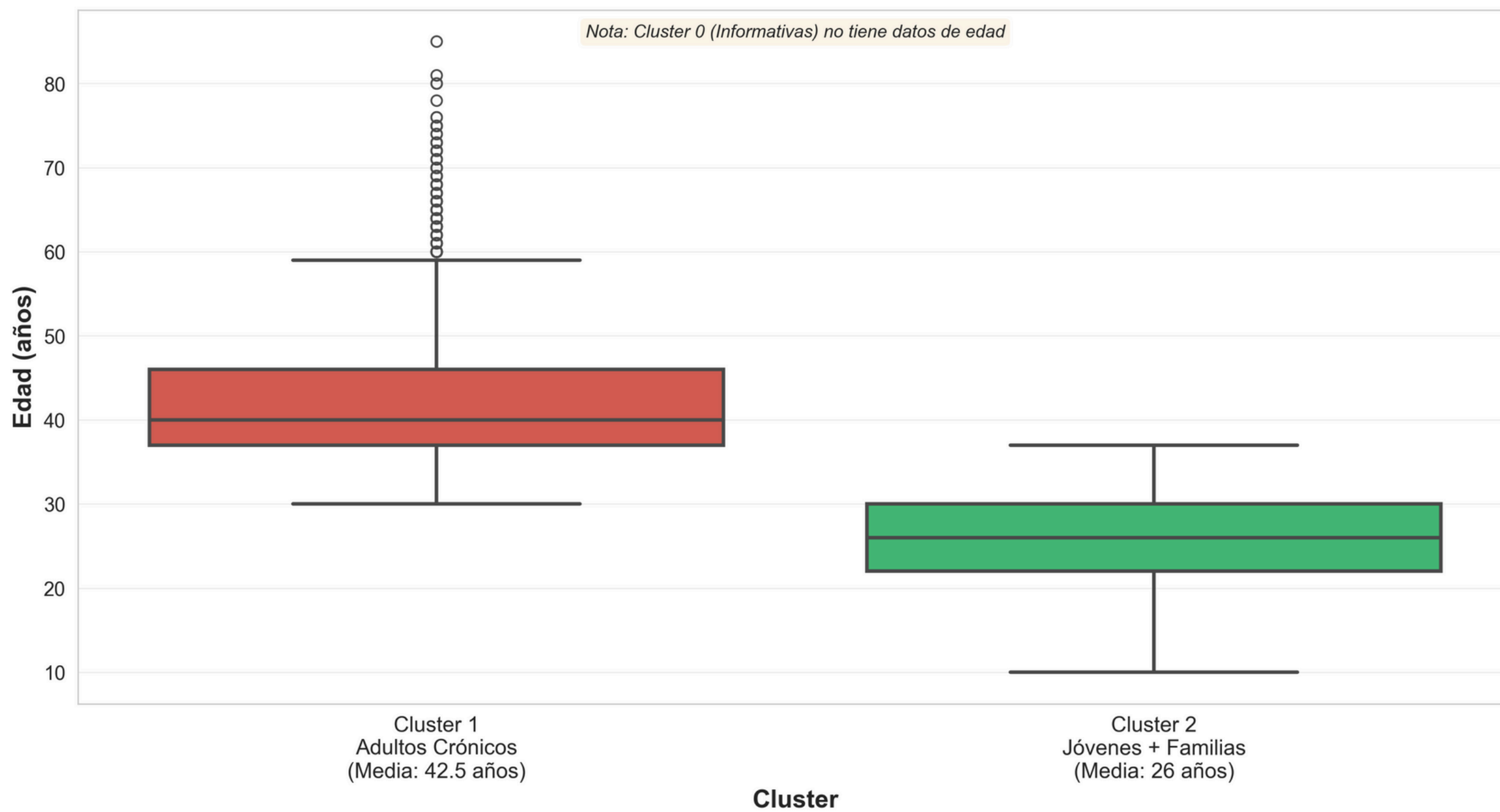
Edad Promedio del Consumidor por Cluster





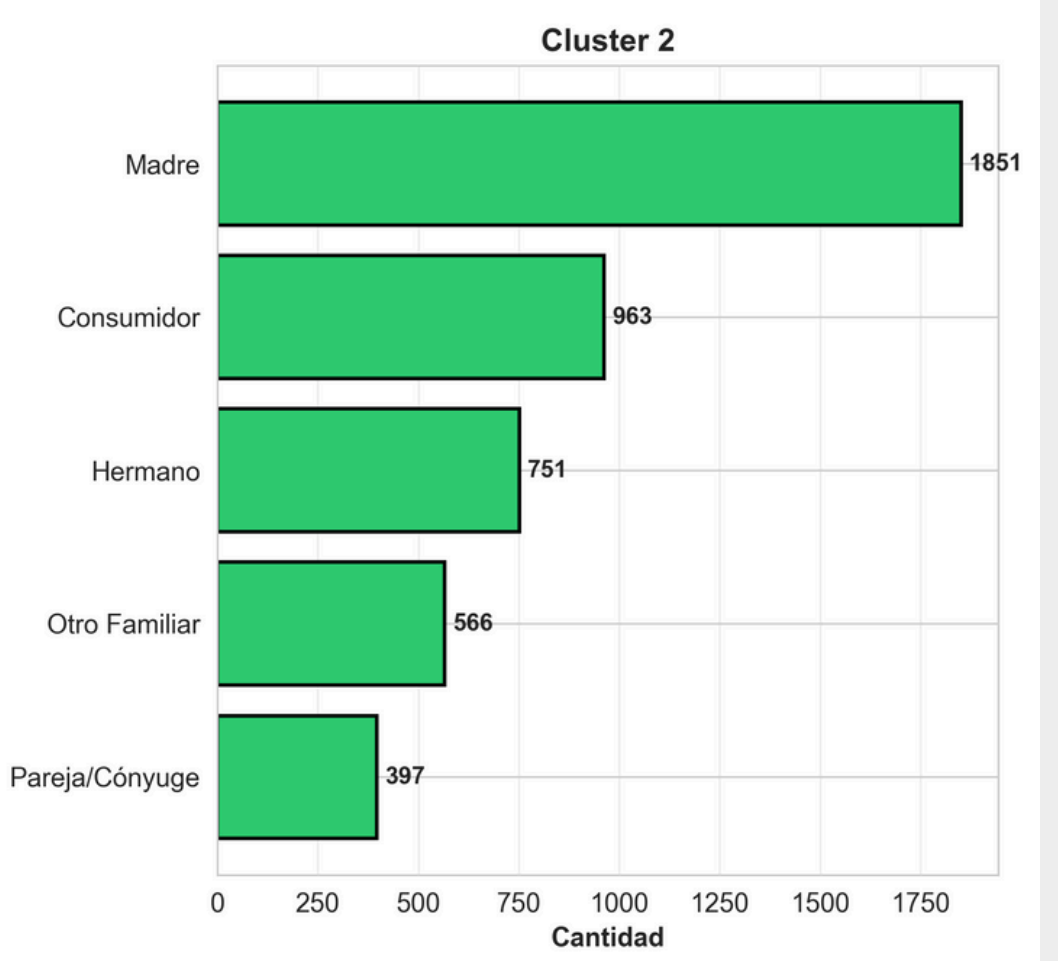
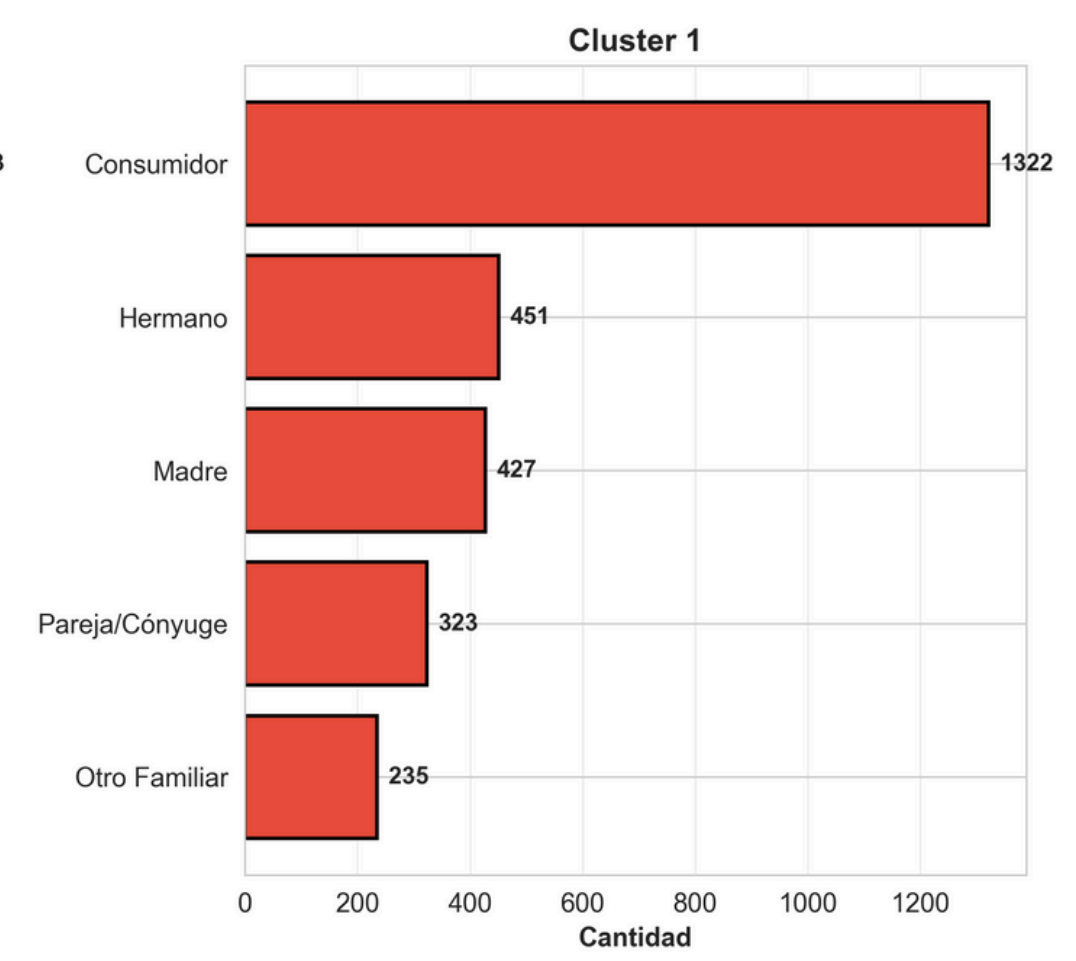
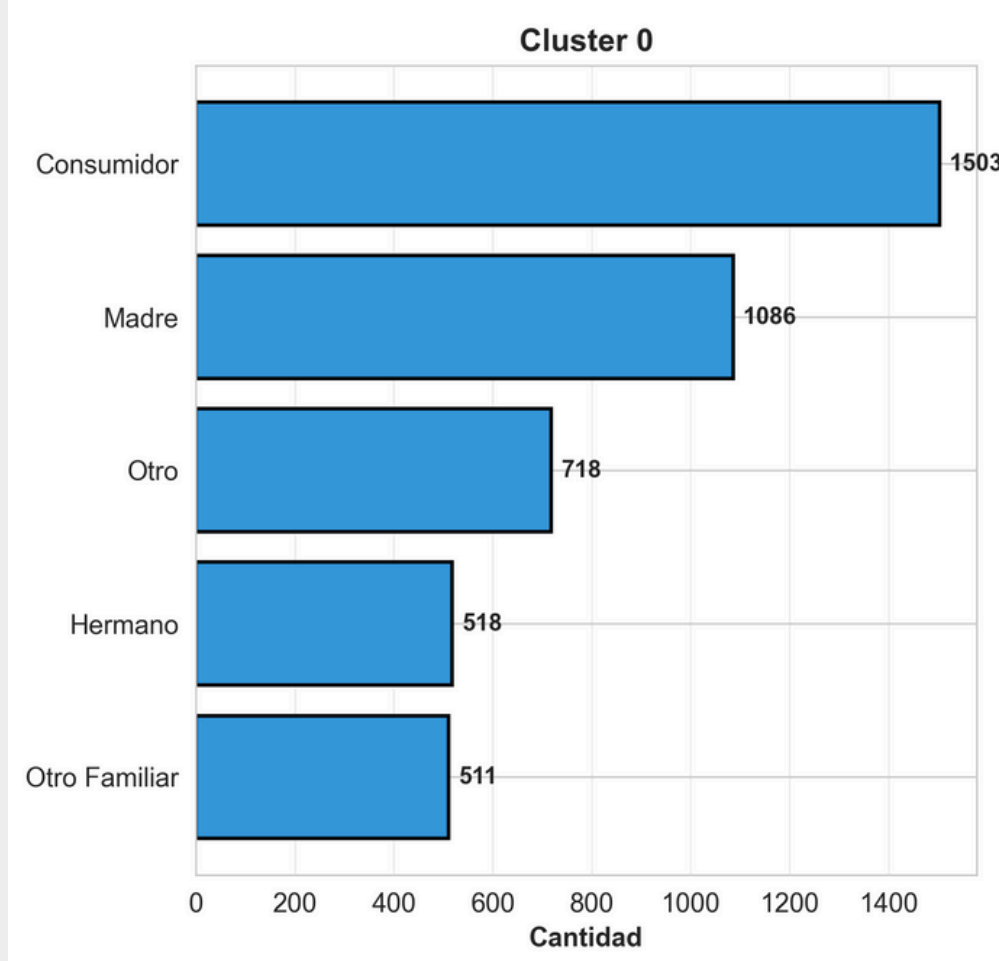


Distribución de Edad del Consumidor por Cluster



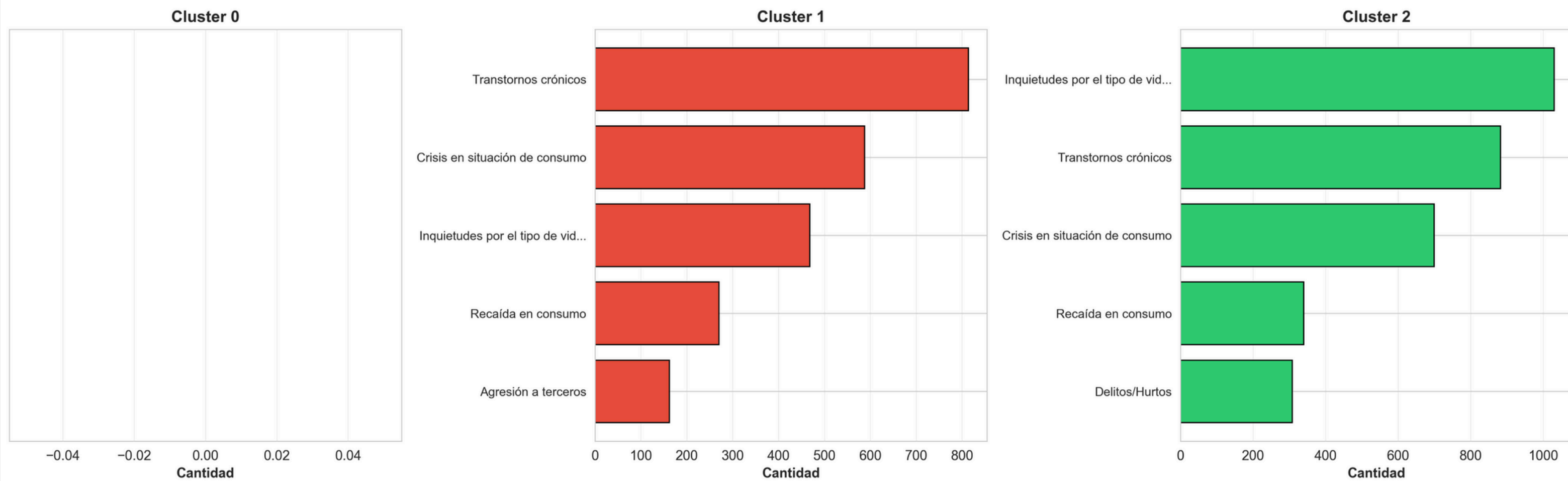


Top 5 Tipos de Consultante por Cluster



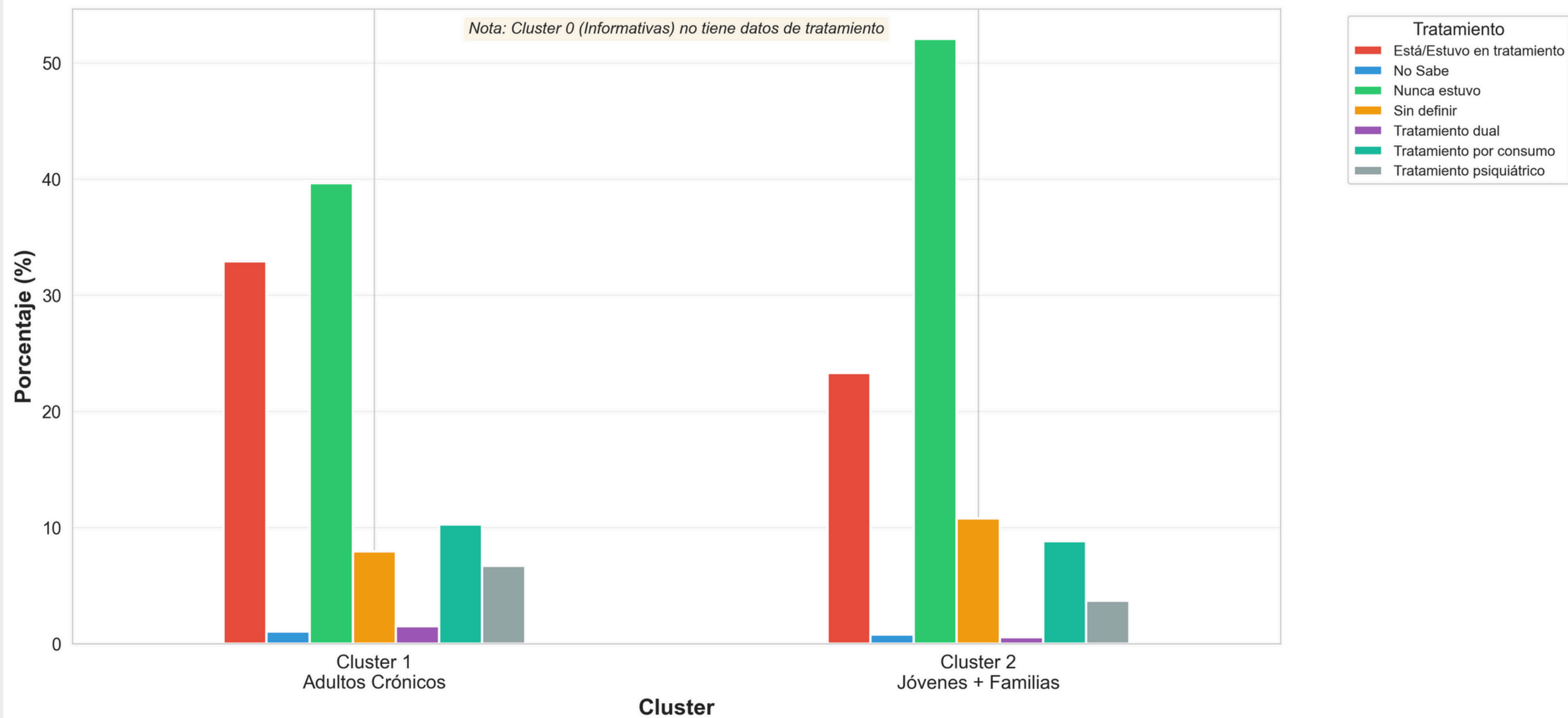


Top 5 Motivos de Consulta por Cluster



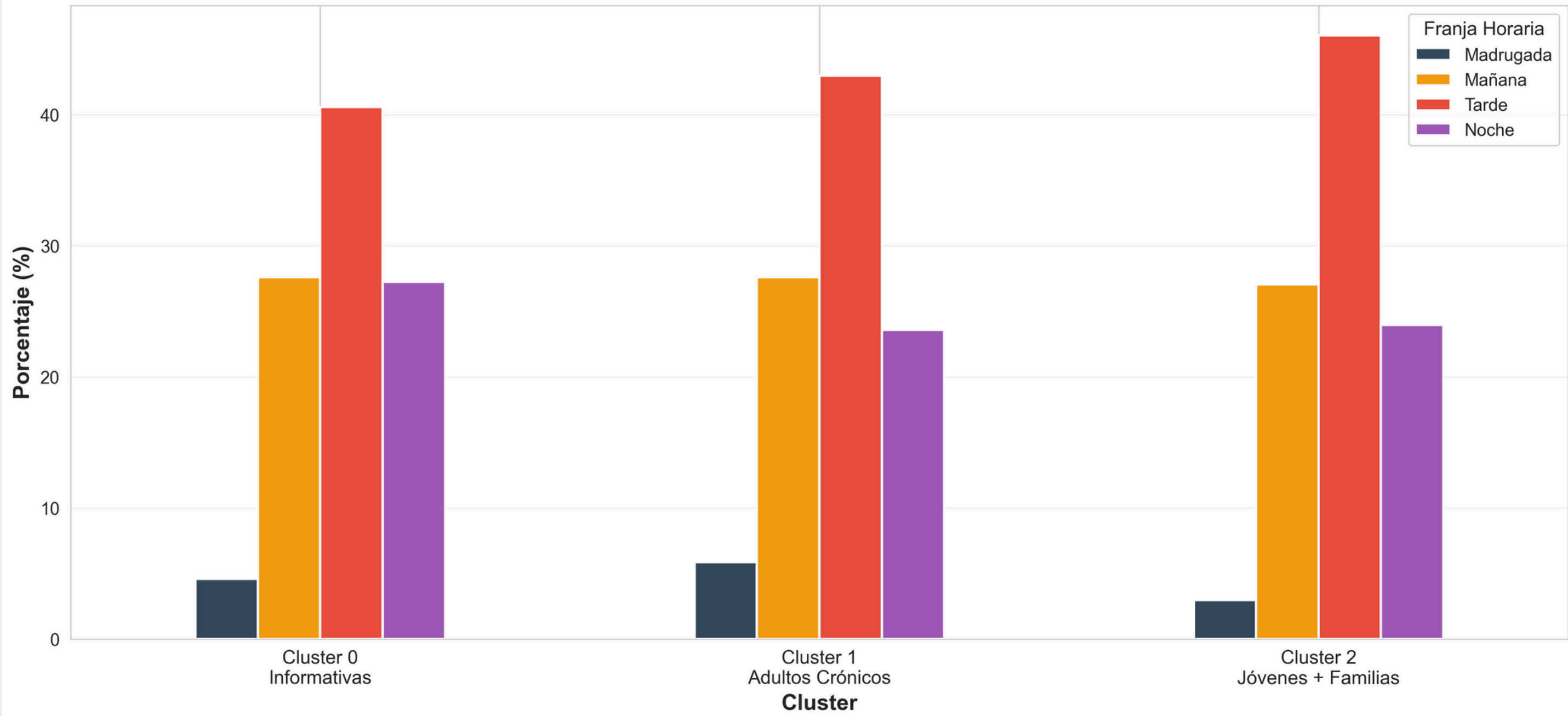


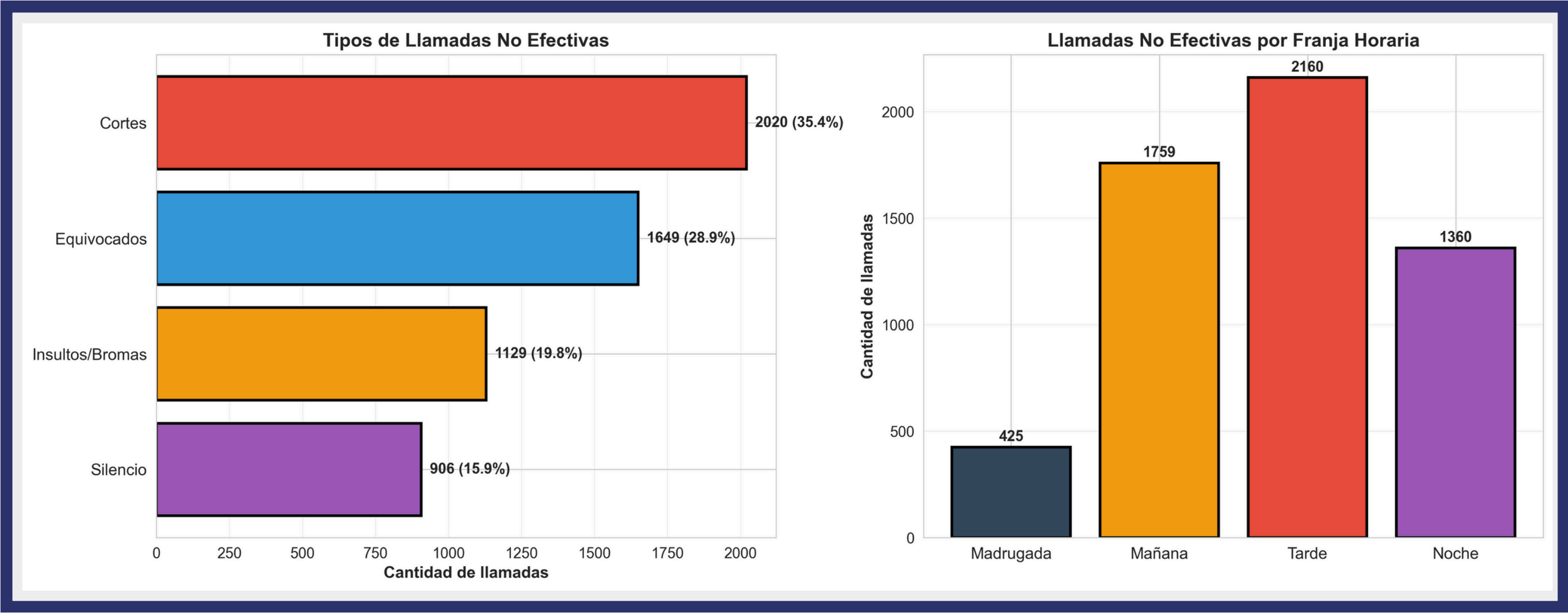
Distribución de Antecedentes de Tratamiento por Cluster





Distribución de Llamadas por Franja Horaria y Cluster








CONCLUSIONES

- El clustering permitió segmentar el universo de llamadas y reconocer perfiles distintos de usuarios.
- Los resultados aportan información estratégica para Sedronar, facilitando la planificación de recursos y campañas de prevención.



CONCLUSIONES

- El modelo es replicable y puede actualizarse con nuevos datos cada semestre para monitorear la evolución del consumo y la demanda de asistencia.
 - Este enfoque promueve una toma de decisiones basada en evidencia, reforzando la eficiencia del servicio público.
- 



**MUCHAS
GRACIAS**

