

Playbook Global · Análisis Portal Ortodoncia

Santiago Tupper J.

[linkedin.com/in/santupper](https://www.linkedin.com/in/santupper)



1) Introducción

Portal Ortodoncia reúne una amplia base de pacientes con distintos niveles de actividad. El desafío que nos planteamos fue simple pero potente: **¿cómo podemos identificar grupos de pacientes para reactivarlos y acompañarlos mejor en sus tratamientos?**

En otras palabras, queríamos pasar de mirar a los pacientes como un "todo" a entenderlos como **segmentos con comportamientos y necesidades distintas**.

👉 Este playbook es el recorrido de ese análisis: desde los datos en bruto hasta los insights de negocio.

2) El viaje de los datos

Trabajamos con varias fuentes internas, siendo la principal el **archivo de clientes**. Allí teníamos información como: edad, montos de presupuestos, recencia de visitas, convenios con empresas, etc.

Para construir un dataset sólido, aplicamos filtros básicos:

- Solo pacientes con **alguna actividad registrada** en presupuestos o atenciones recientes.
- Variables limpias, sin duplicados ni ruido excesivo.
- Transformación de categorías (por ejemplo, si un paciente pertenecía a una empresa, lo convertimos en un indicador).

El resultado fue un conjunto de datos listo para analizar, equilibrado entre lo clínico y lo comercial.

Edad	Sexo	FechaRegistro	ClinicaCreaPac	ISAPRE	Empresa	EsTitular	CantPptos	TotPptos	CantPptosAbo	TotPptosAbo	CantPptosAvan	TotPptosAvan	Atencion15d	Atencion1m
46	F	2021-01-15 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
59	F	2021-12-13 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
60	F	2022-04-20 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
42	F	2022-01-05 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
45	F	2023-02-09 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
58	F	2020-02-14 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
44	F	2021-12-02 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
57	F	2020-11-17 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
44	F	2022-09-28 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
56	F	2021-06-07 00:00:00.000	Providencia	NaN	NaN	TITULAR	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

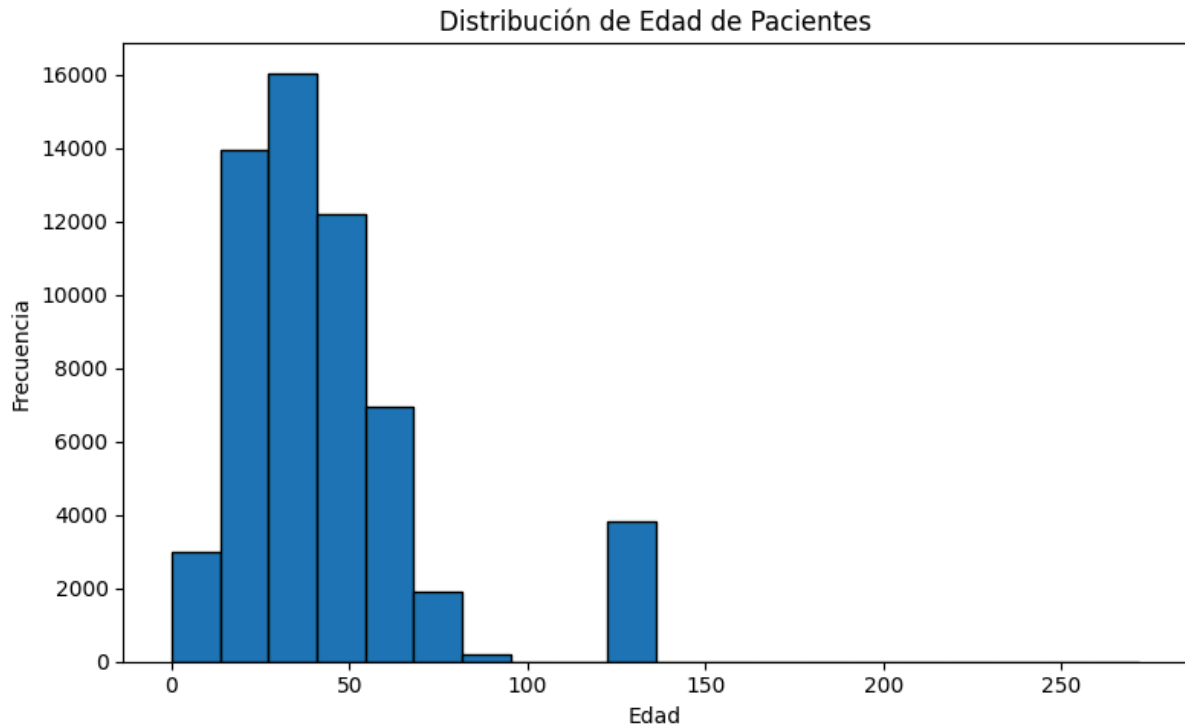
3) Explorando los datos (EDA)

Antes de modelar, exploramos para entender la “personalidad” de los pacientes.

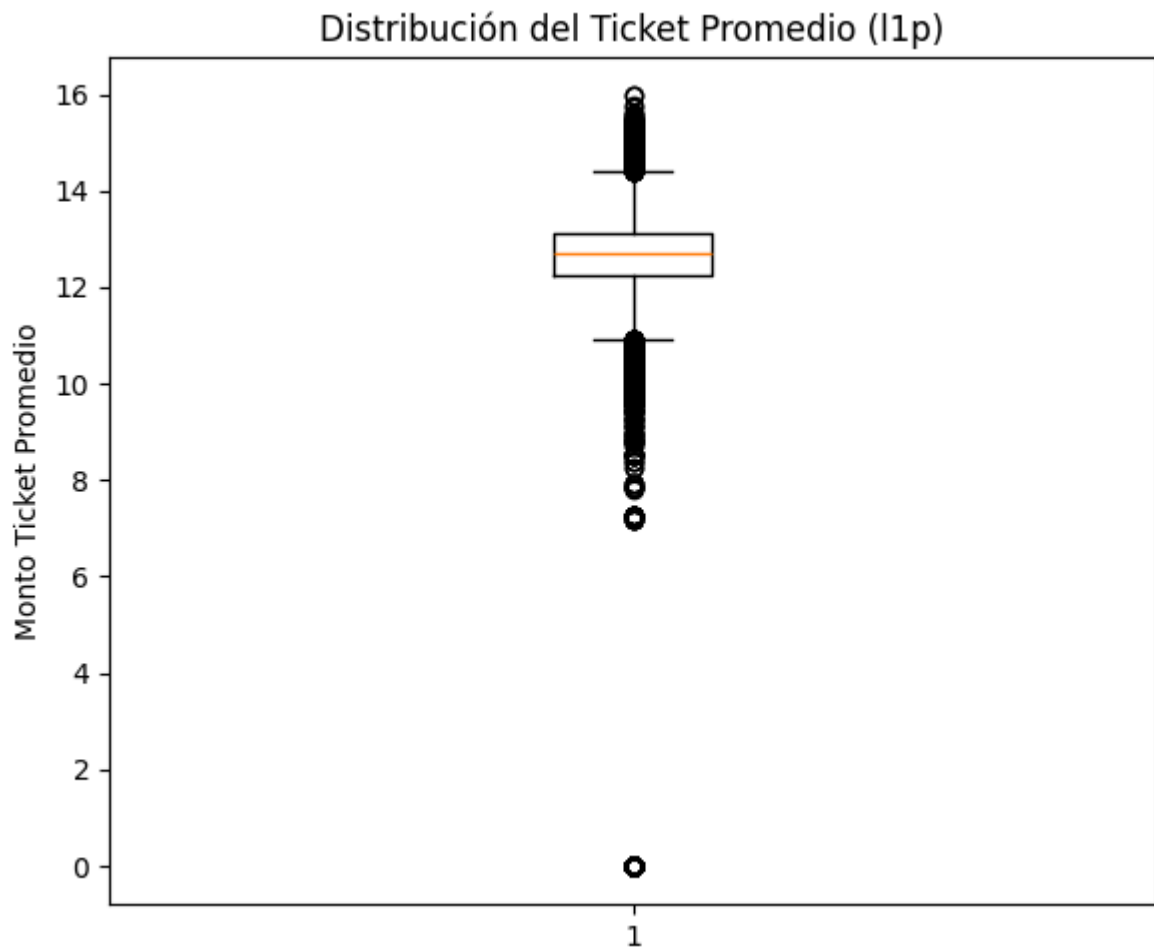
Algunas observaciones clave:

- La **edad** está distribuida de manera bastante amplia, con grupos jóvenes y adultos mayores.
- El **ticket promedio** varía mucho: algunos pacientes concentran montos altos, pero la mayoría tiene valores modestos.
- La **recencia de atención** (visitas en 15 días, 1 mes, 3 meses) muestra que muchos pacientes tienen largos periodos sin regresar.

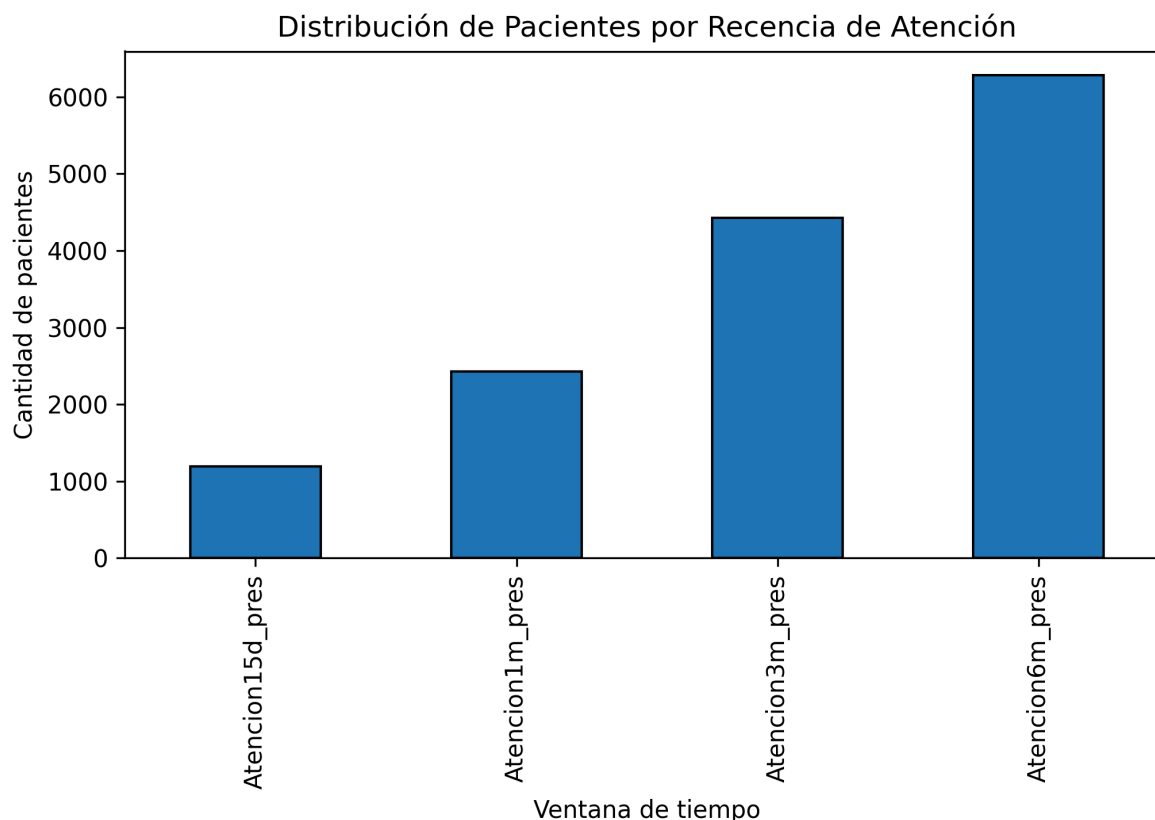
👉 Esta exploración fue la base para decidir cómo escalar los datos y qué variables usar en los grupos.



Refleja la concentración de pacientes en rangos etarios medios, clave para segmentar la comunicación.



Muestra la dispersión del gasto por paciente, con valores extremos que elevan el promedio.

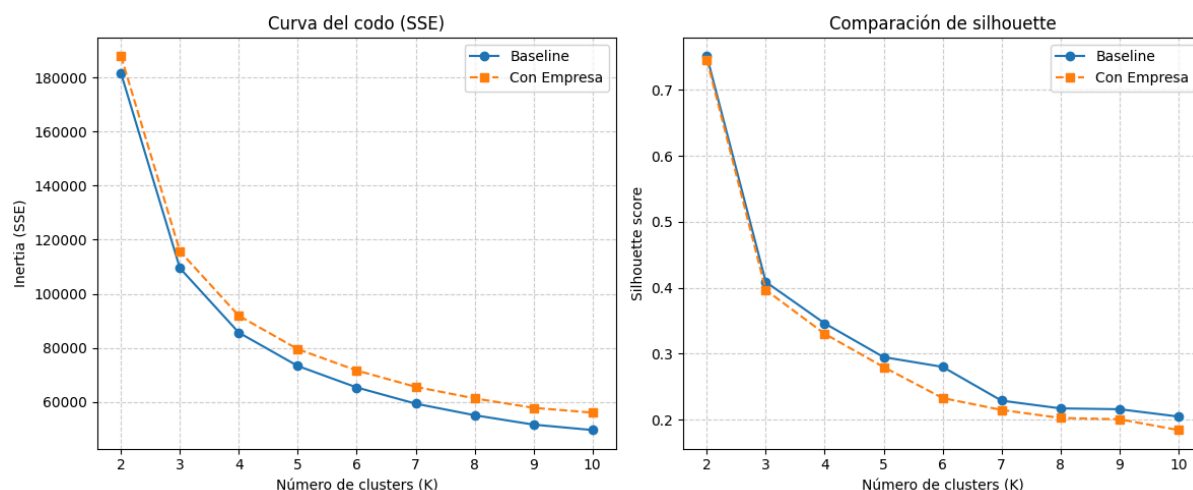


Evidencia cuántos pacientes llevan meses sin visitar la clínica, mostrando oportunidades de reactivación.

4) Construyendo los grupos (Clustering)

El corazón del análisis fue aplicar un algoritmo de clustering. La idea: **dejar que los datos nos digan qué grupos existen** sin imponer categorías previas.

Para decidir cuántos grupos usar, probamos distintas opciones y analizamos cuál daba la mejor separación y coherencia. Finalmente, elegimos un número de clusters que equilibraba simplicidad con riqueza de insights.



Permite determinar la cantidad óptima de clusters, combinando criterio geométrico y calidad de segmentación.

5) Conociendo a nuestros clusters

Aquí es donde el proyecto cobra vida: traducimos números en **perfiles de pacientes**.

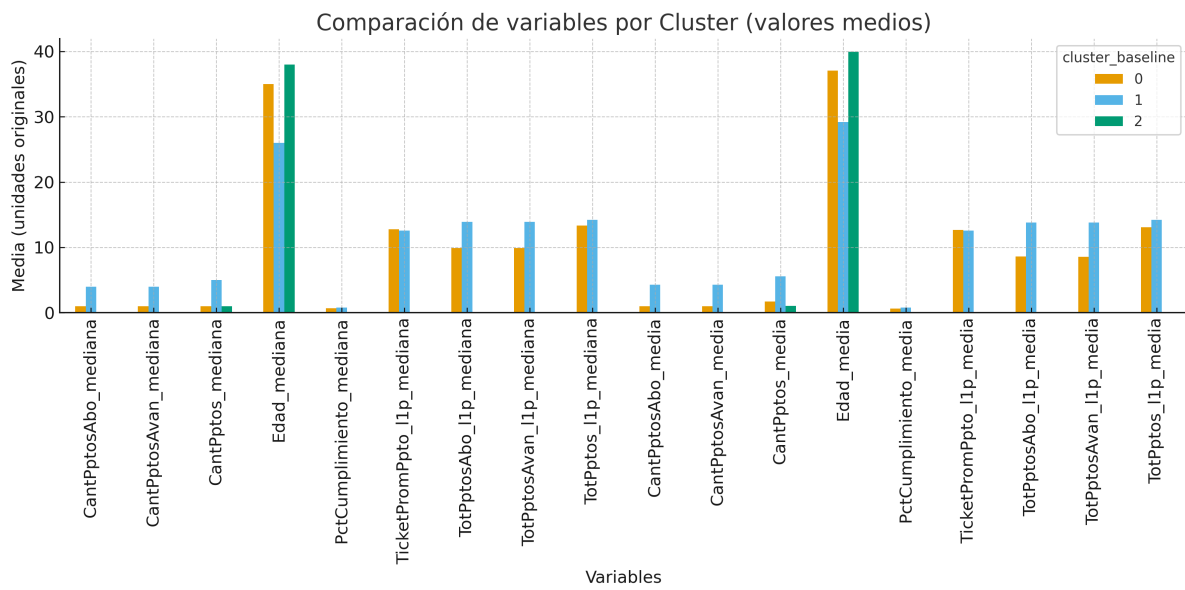
Ejemplos de lo que encontramos:

- **Recurrentes – Alto Ticket:** pacientes que vienen seguido y generan montos altos. Ideales para programas de fidelización premium.
- **Dormidos – Bajo Ticket:** llevan tiempo sin venir, con montos bajos. Claros candidatos a campañas de reactivación con incentivos.
- **Ocasionales – Ticket Medio:** pacientes intermitentes, que necesitan recordatorios y paquetes que fomenten continuidad.

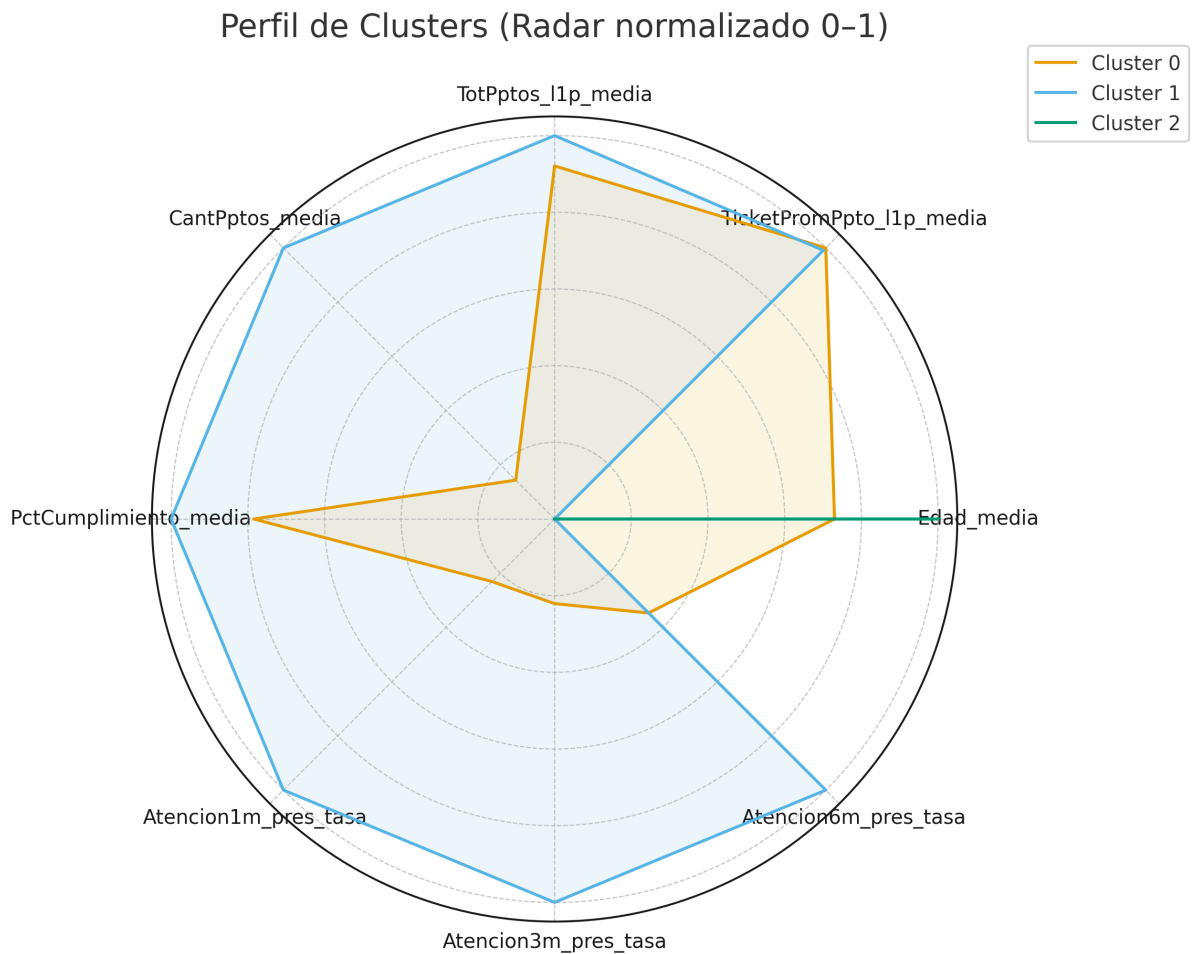
Cada cluster tiene su propia historia y nos permite diseñar estrategias personalizadas.

cluster_baseline	Pacientes	Edad (mediana)	Presupuestos (mediana)	Monto total (log1p, mediana)	% Cumplimiento (mediana)	Presencia 6m (%)	Presencia 3m (%)
0	10613	35.0	1.0	13.315	0.667	40.2	26.3
1	3216	26.0	5.0	14.203	0.8	59.8	48.9
2	312	38.0	1.0	0.0	0.0	29.8	19.9

Resume las principales variables promedio de cada cluster, funcionando como perfil ejecutivo.



Destaca cómo difieren los clusters en variables clave, facilitando la interpretación práctica.



Visualiza de forma comparativa las fortalezas y debilidades de cada cluster en múltiples dimensiones.

6) Validando que los grupos tengan sentido

No queríamos que los clusters fueran solo un resultado matemático. Por eso, los validamos contra otras fuentes:

- **Power BI:** chequeamos que los totales y distribuciones coincidieran con los reportes oficiales de la clínica.
- **Prestaciones:** revisamos si ciertos clusters consumían tipos específicos de tratamientos.

El resultado fue consistente: los grupos reflejaban **patrones reales de comportamiento clínico**.

Validación con Power BI

Para asegurar consistencia global:

- Total de pacientes: Dashboard (≈14.500) vs. Clustering (14.141).
- Diferencias mínimas explicadas por corte temporal (PBI incluye un mes posterior).
- % con atención en 6m, % cumplimiento y % RM ≈ idénticos en ambos.

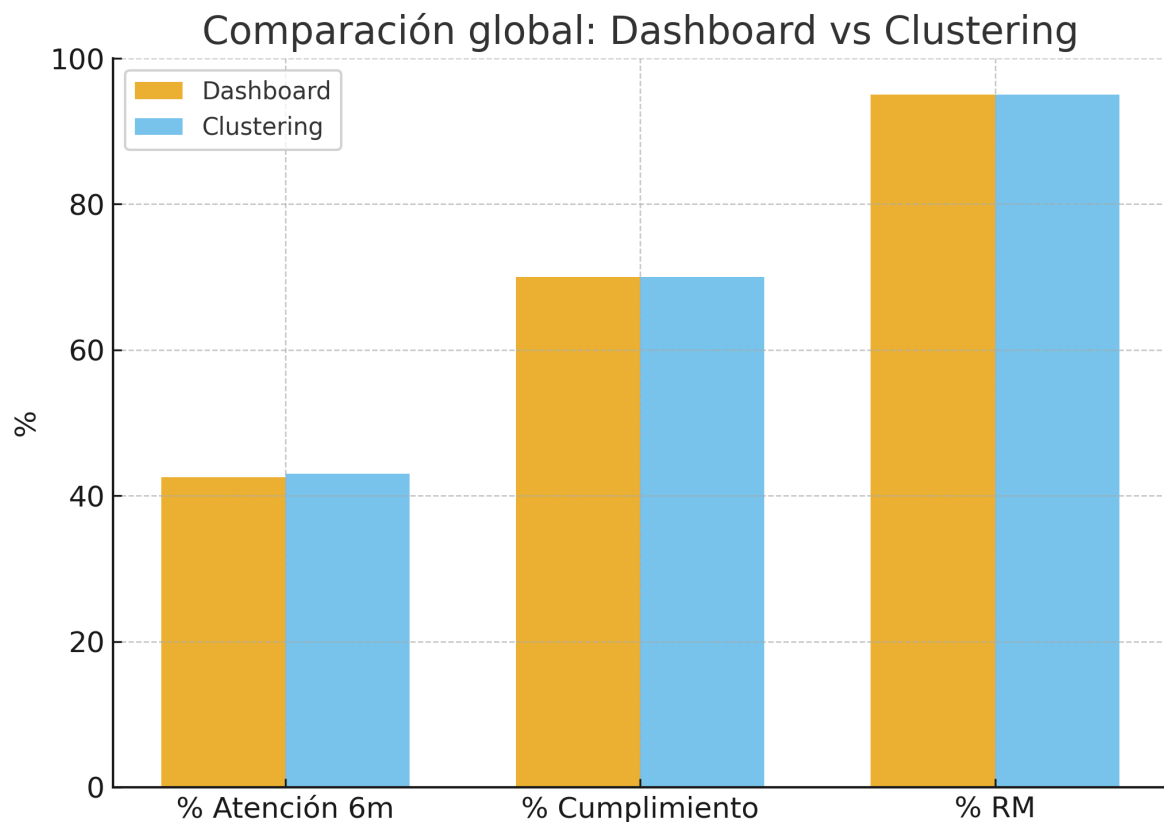
Evidencias visuales:



Muestra los indicadores globales oficiales de la clínica para validar los resultados del clustering.

Metrica	Dashboard (≈jul/2025)	Clustering (jun/2025)
Pacientes activos	14500 aprox	14141
% con atención en 6m	40–45%	Cluster 0: 40% Cluster 1: 60% Cluster 2: 30%
% Cumplimiento	≈70% global	Cluster 0: 67% Cluster 1: 80% Cluster 2: 0%
Distribución geográfica (RM)	≈95% RM / 5% otras	Cluster 0: 97,5% RM Cluster 1: 98,3% RM Cluster 2: 61% RM
Empresa/Convenio	Fonasa mayoritario / Isapres 20–25%	Cluster 0: Fonasa 75%, Isapres 9% Cluster 1: Fonasa 69%, Isapres 26% Cluster 2: mezcla atípica

Confirma la consistencia entre métricas globales del clustering y el dashboard interno.



Contrasta visualmente los promedios del clustering con los del dashboard para validar diferencias menores.

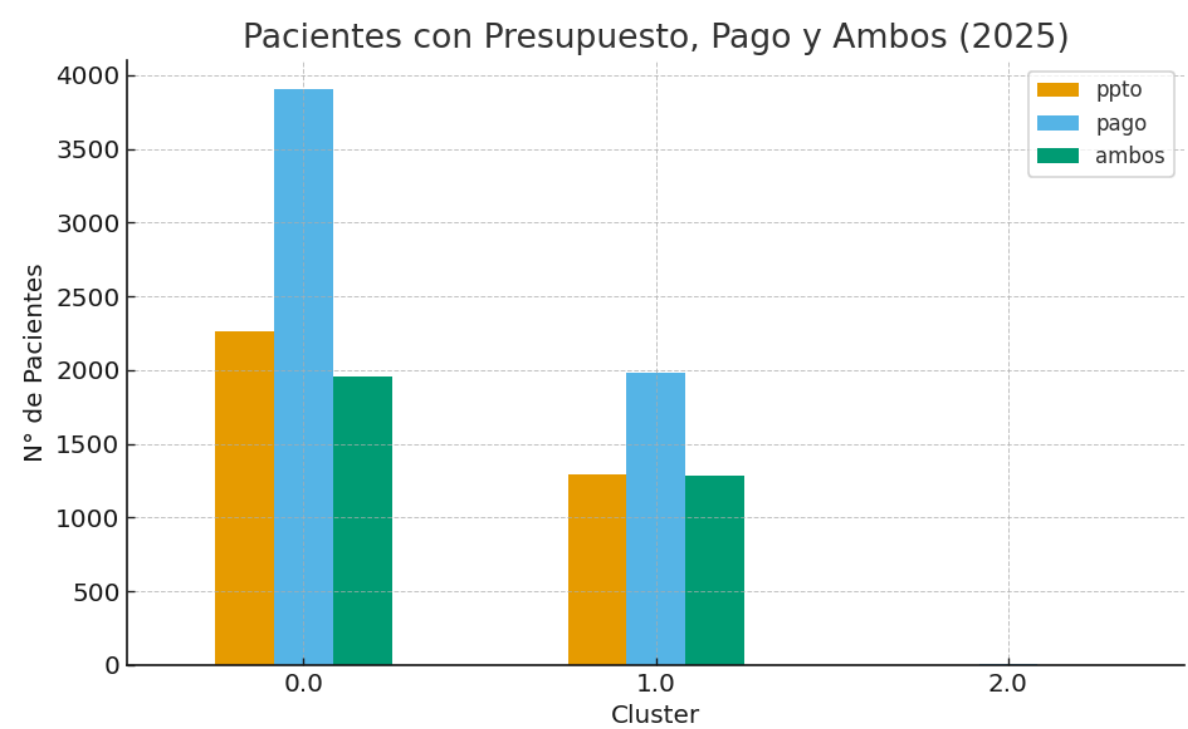
🦷 Validación con Prestaciones

Analizamos el cruce entre clusters y tipos de prestaciones:

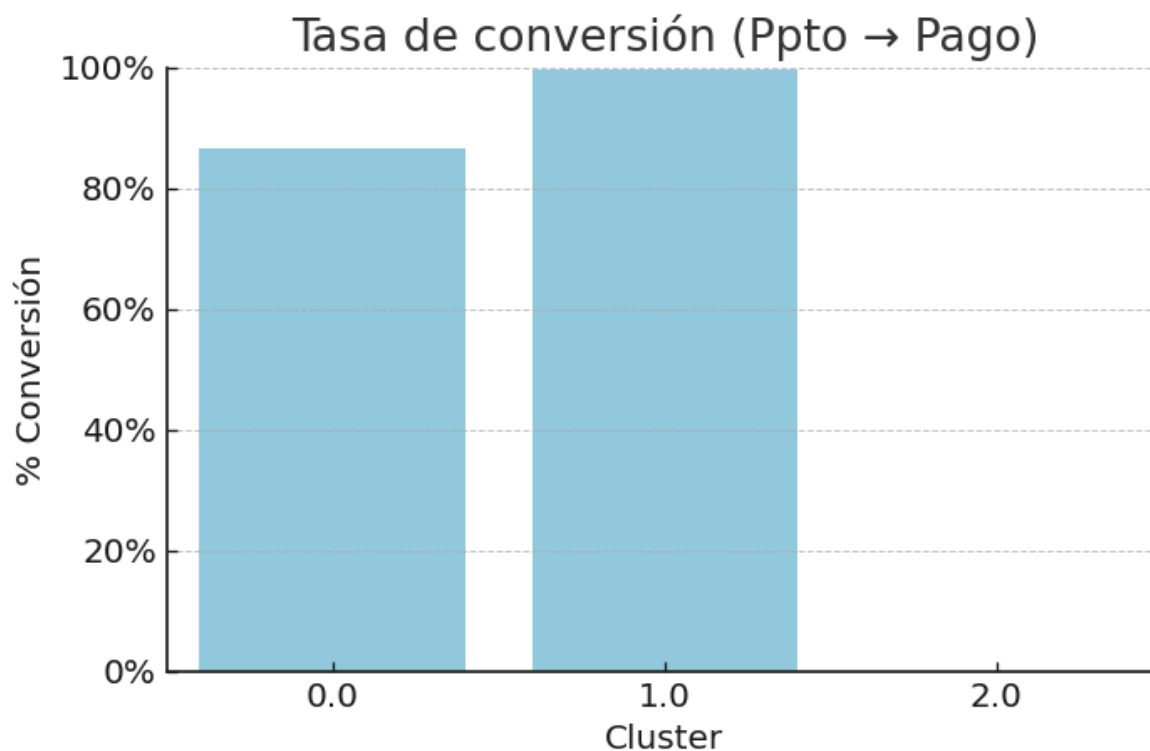
- Cada cluster consume combinaciones distintas de tratamientos.
- Ejemplo: Cluster 1 concentra prestaciones de seguimiento, Cluster 2 explica mayor dispersión fuera de RM.
- Esto confirma que los clusters reflejan **patrones clínicos reales**, no solo artefactos del modelado.

Evidencias visuales:

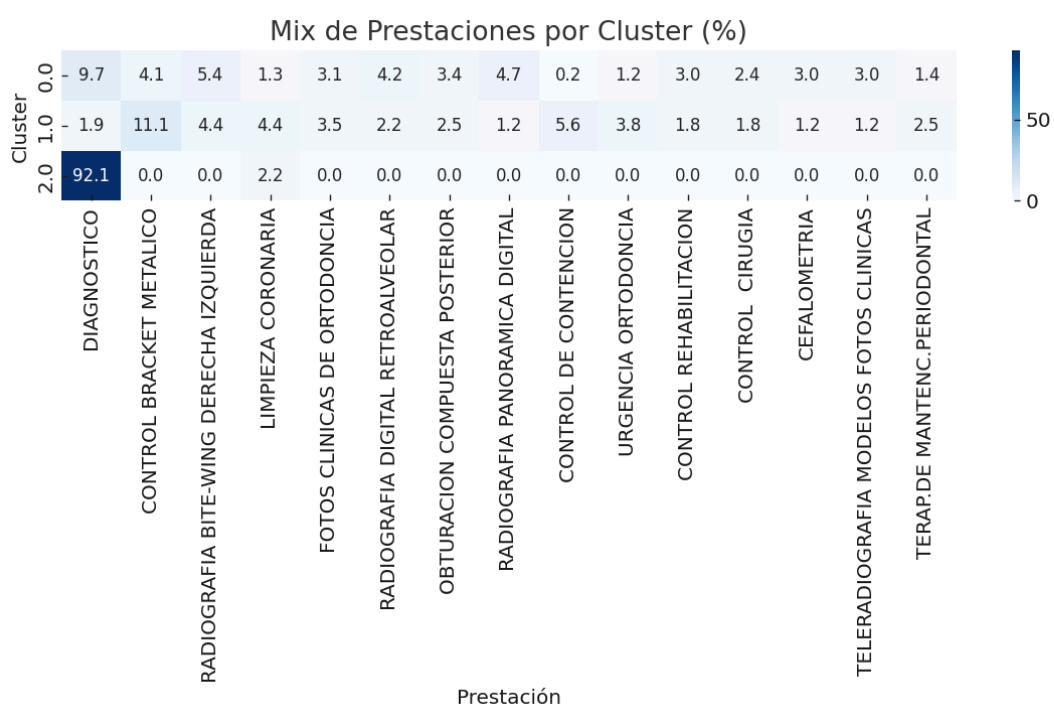
Corte	Pacientes_con_Ppto	Pacientes_con_Pago	Pacientes_con_Ambos	Tasa_Conversion
2025	4564	4862	3951	0.8656879929886064



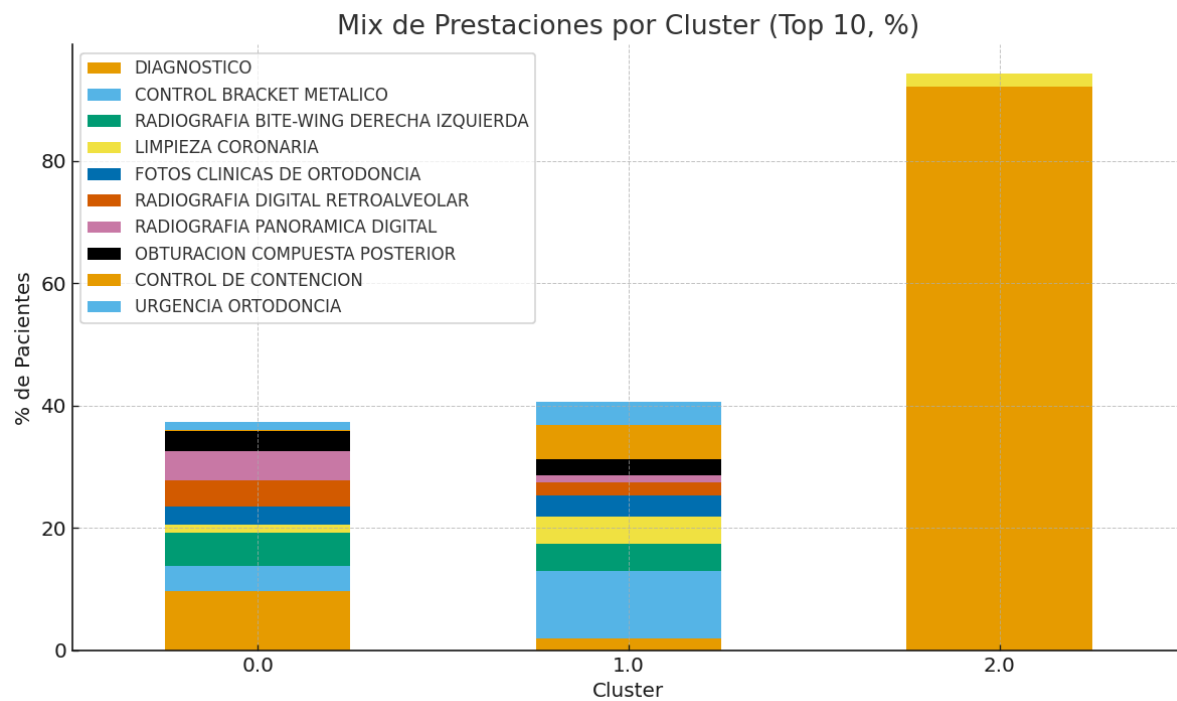
Comparación de pacientes con presupuesto, pago y ambos en 2025, desglosados por cluster.



Tasa de conversión (pacientes con presupuesto que terminaron pagando) por cluster.



Distribución porcentual de las principales prestaciones dentro de cada cluster (heatmap).

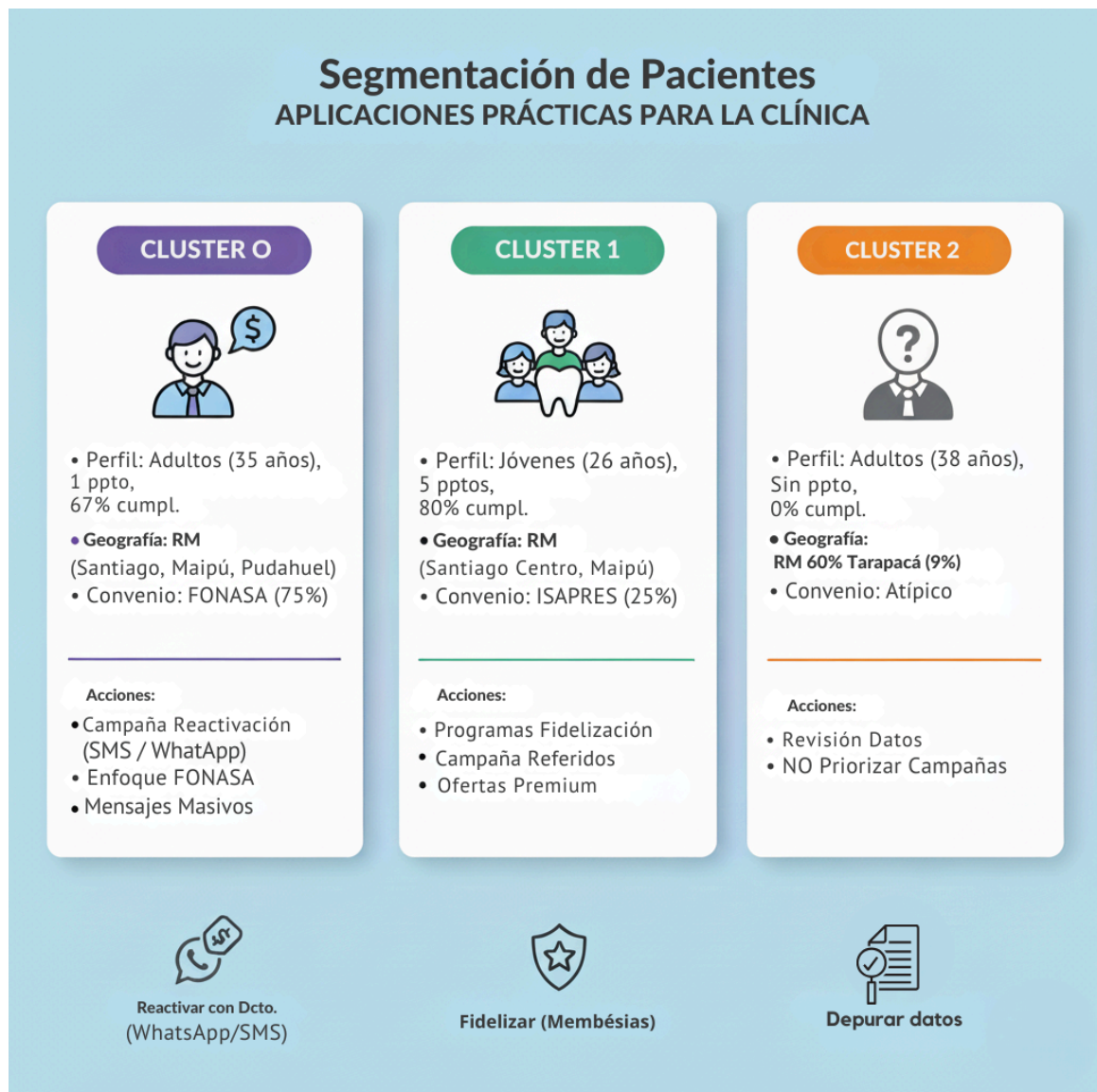


Mix de las 10 prestaciones más frecuentes, representado en porcentajes con barras apiladas por cluster.

7) Aplicaciones prácticas para la clínica

Los insights no se quedan en teoría: abren la puerta a estrategias concretas.

- **Reactivar pacientes dormidos** con campañas simples por WhatsApp o SMS, ofreciendo descuentos de bienvenida.
- **Fidelizar a los recurrentes** con membresías o chequeos estéticos adicionales.
- **Optimizar la operación** priorizando llamadas y esfuerzos comerciales en segmentos de mayor retorno.
- **Revisar y depurar datos** para tener una base consistente.



8) Lecciones aprendidas

El proyecto dejó aprendizajes valiosos tanto técnicos como de gestión del proceso:

- **Metodología técnica**

- Los outliers distorsionan resultados si no se controlan: usar **escaladores robustos o mixtos** si es necesario y validar siempre las transformaciones es esencial.
- La validación cruzada con fuentes externas refuerza la confianza en los resultados y muestra el valor agregado del clustering.

- **Estructura y organización**

- Aprendí a **organizar un proyecto de análisis en carpetas estándar** (`data/raw`, `data/processed`, `notebooks`, `reports`, etc.), lo que hace más fácil mantener orden y trazabilidad.
- Entendí la importancia de usar `.gitignore` para que los archivos sensibles o demasiado pesados no se suban al repositorio.
- GitHub funciona como repositorio **técnico**, mientras que Notion cumple el rol narrativo y visual: juntos hacen posible comunicar tanto a reclutadores como a directivos no técnicos.

- **Proceso de trabajo**

- No basta con obtener métricas: hay que **traducir resultados en storytelling**, con imágenes, captions y callouts que transmitan el valor a quienes no son técnicos.
- El **worklog diario** resultó ser una herramienta muy útil para documentar avances, decisiones y pendientes de forma continua.
- El aprendizaje no fue solo de Python o modelos, sino de cómo **documentar y versionar un proyecto de punta a punta**, con disciplina y claridad.



“No basta con analizar; hay que contar la historia de manera que otros la entiendan”.

9) Conclusiones

El proyecto no solo permitió segmentar pacientes y descubrir patrones clínicos relevantes, sino también sentar las bases de un **proceso de análisis reproducible y documentado**.

- Los clusters obtenidos mostraron perfiles diferenciados con **implicancias prácticas directas** en comunicación, fidelización y gestión comercial.
- La validación contra fuentes oficiales aseguró que los resultados fueran consistentes y confiables.
- La combinación de **herramientas técnicas (Python, GitHub)** con **herramientas narrativas (Notion, infografías, storytelling)** demostró que los análisis pueden traducirse en propuestas claras y accionables.

En síntesis: este trabajo dejó no solo un resultado analítico, sino también una metodología replicable para futuros proyectos de la clínica o de otros contextos.