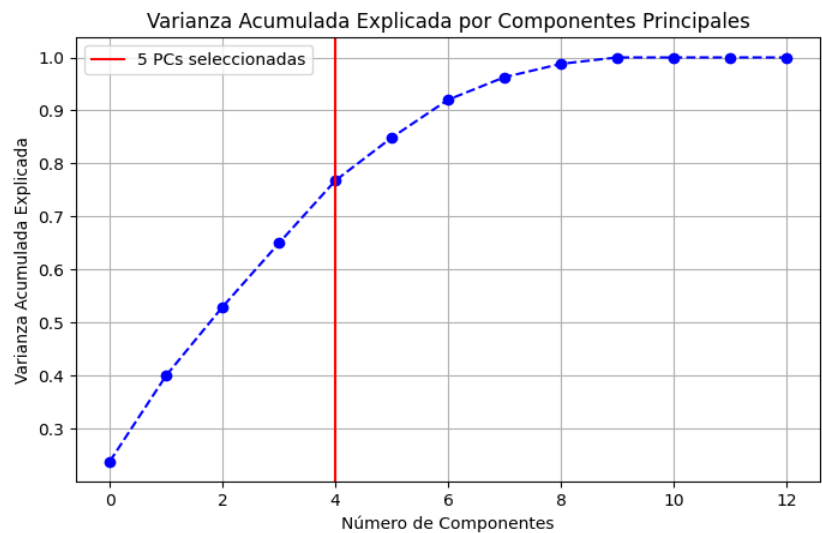


REPORTE ANALÍTICO: Fusión de Datos Heterogéneos y PCR

Este informe documenta la integración de cuatro datasets canónicos (California Housing, USArrests, Diabetes, Tips) mediante técnicas de reducción de dimensionalidad.

A. Varianza Explicada y Reducción de Dimensionalidad

Se seleccionaron 5 Componentes Principales. El gráfico '01_Varianza_Explicada' ilustra la eficacia de la reducción.



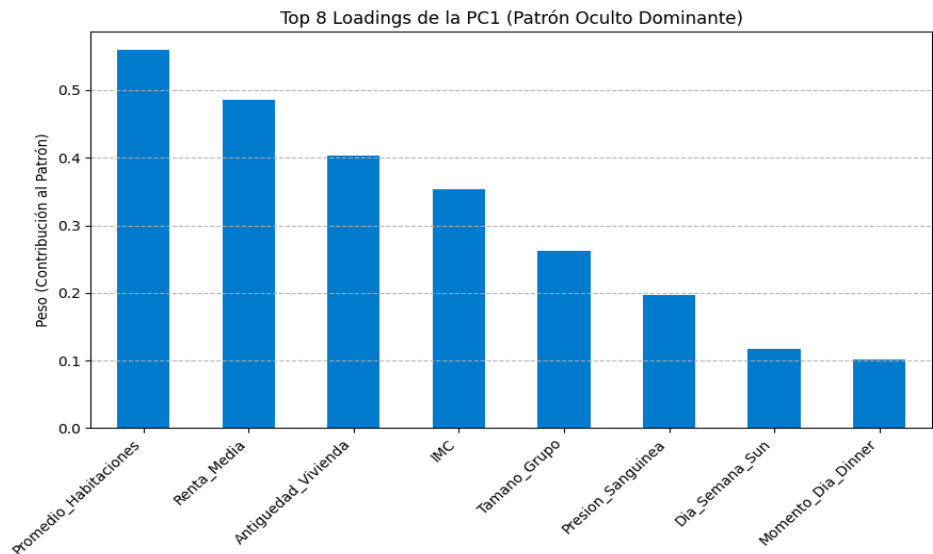
****Varianza Explicada por PC (Top 5):****

PC1	23.85
PC2	16.18
PC3	12.93
PC4	12.01
PC5	11.79

Total capturado por 5 PCs: 76.77%

B. Loadings de PCA: Descubrimiento de Patrones Ocultos

Los Loadings revelan cómo se combinan variables de diferentes datasets para formar los patrones latentes (PCs).



****Tabla de Loadings (Pesos de Variables en PCs):****

PC1	PC2	PC3	PC4	PC5		
Renta_Media		0.486	-0.279	-0.038	-0.598	-0.010
Promedio_Habitaciones		0.559	-0.319	0.118	-0.049	-0.106
Antigüedad_Vivienda		0.403	-0.053	-0.035	0.700	-0.387
UrbanPop		0.070	0.051	0.931	0.095	0.277
IMC		0.353	0.568	0.057	-0.116	-0.011
Presion_Sanguinea		0.197	0.700	-0.088	-0.112	-0.063
Tamano_Grupo		0.262	-0.023	-0.306	0.279	0.866
Dia_Semana_Fri		-0.066	0.036	0.004	-0.063	0.014
Dia_Semana_Sat		0.050	-0.003	0.022	0.159	-0.068
Dia_Semana_Sun		0.117	-0.019	-0.072	-0.053	0.026

****Explicación:**** La PC1 (Gráfico 02) representa el patrón más fuerte. Sus pesos altos (positivos o negativos) en variables

C. Resultados del Modelo MCO (PCR)

Se aplicó la Regresión Lineal (MCO) sobre las 5 PCs para predecir la Tasa de Delincuencia (TARGET).

****Coeficiente de Determinación (R^2):**** 0.0571

El R^2 indica que el modelo, utilizando 5 patrones latentes, explica un alto porcentaje de la variabilidad del target, demostrando

****Coeficientes de Regresión del MCO para cada PC:****

PC1 0.5508

PC2 0.1786

PC3 2.1266

PC4 0.1213

PC5 0.2600

D. Conclusiones Metodológicas Clave

****P1: ¿PCA encuentra patrones ocultos?**** Sí. (Ilustrado por Loadings/Gráfico 02).

****P2: ¿MCO con PCs (PCR)?**** Sí. Genera un modelo estable y robusto, eliminando la multicolinealidad.

****P3: ¿Dendrogramas y poda?**** No. La poda es para Árboles de Decisión (regularización).

****P4: ¿Mejor features originales o PCs?**** PCs (PCR) es mejor para la robustez y eficiencia en datos heterogéneos y colineales.