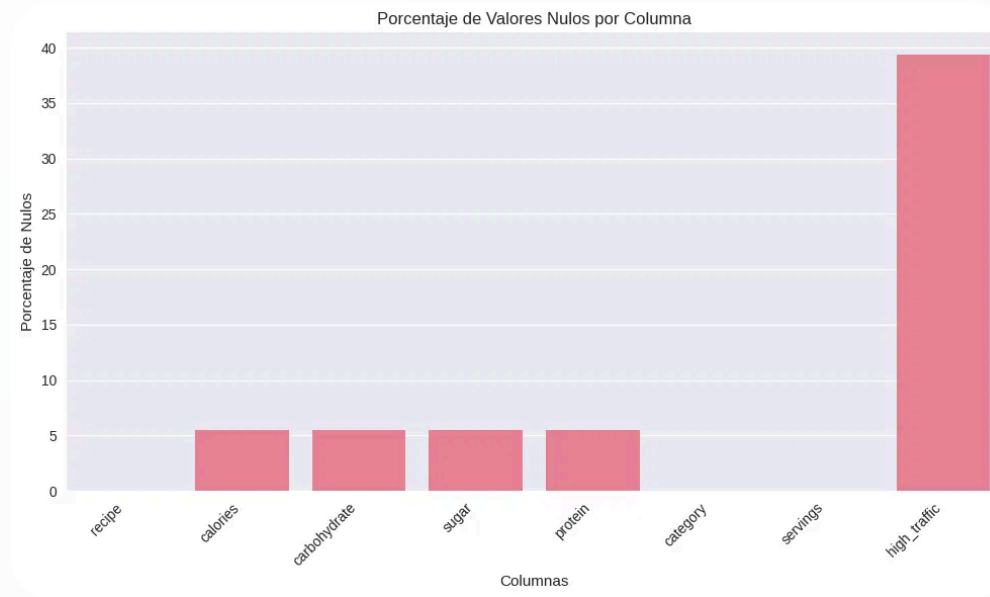


# Análisis de Clustering de Recetas

## Patrones Nutricionales en Recetas



Hellen Carranza

# Introducción y Objetivo



El **análisis de clustering** nos permite descubrir patrones ocultos en las recetas, agrupándolas según sus similitudes nutricionales.



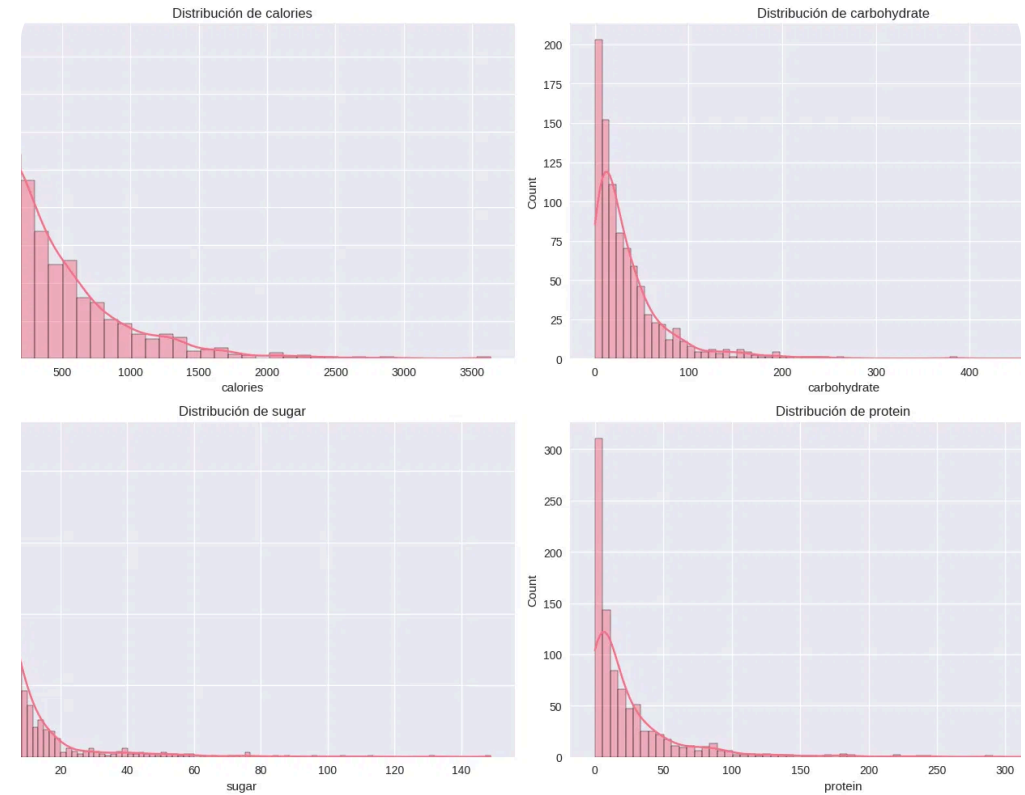
Entender estos patrones nos ayuda a crear recomendaciones personalizadas y a diseñar menús más equilibrados y variados.



**Nuestro objetivo:** Identificar grupos naturales de recetas basados en su perfil nutricional (calorías, carbohidratos, azúcar y proteínas).



Los resultados pueden transformar la forma en que organizamos recetas en sitios web, aplicaciones de nutrición y planificación de dietas.



# Metodología

## 1 Preprocesamiento de Datos

Seleccionamos las variables nutricionales clave y eliminamos las recetas con valores incompletos.

## 2 Transformación Logarítmica

Aplicamos transformación  $\log_{10}$  para suavizar la distribución y reducir el impacto de valores extremos.

## 3 Escalado de Datos

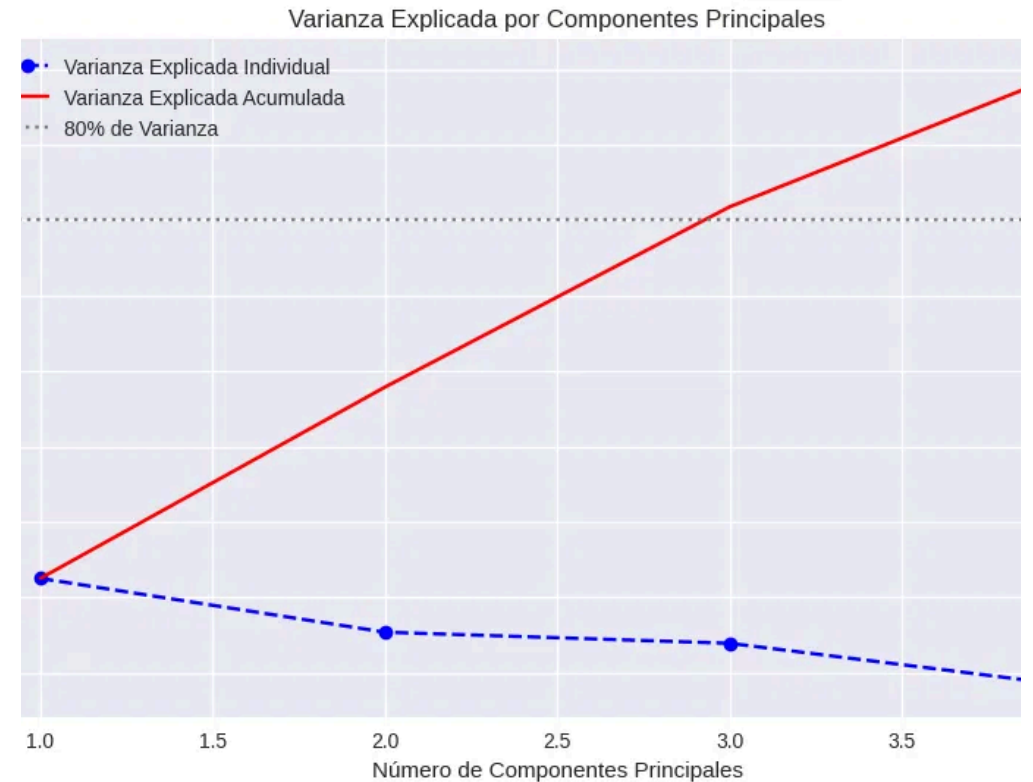
Utilizamos RobustScaler para normalizar las características y hacerlas comparables entre sí.

## 4 Algoritmos de Clustering

Probamos K-Means, Clustering Jerárquico, DBSCAN y Spectral Clustering para encontrar el mejor enfoque.

## 5 Evaluación y Selección

Comparamos los resultados usando métricas como Silhouette Score para elegir el algoritmo óptimo.



# Exploración de Datos


 Dataset Original

947 recetas

 Dataset Limpio


895 recetas

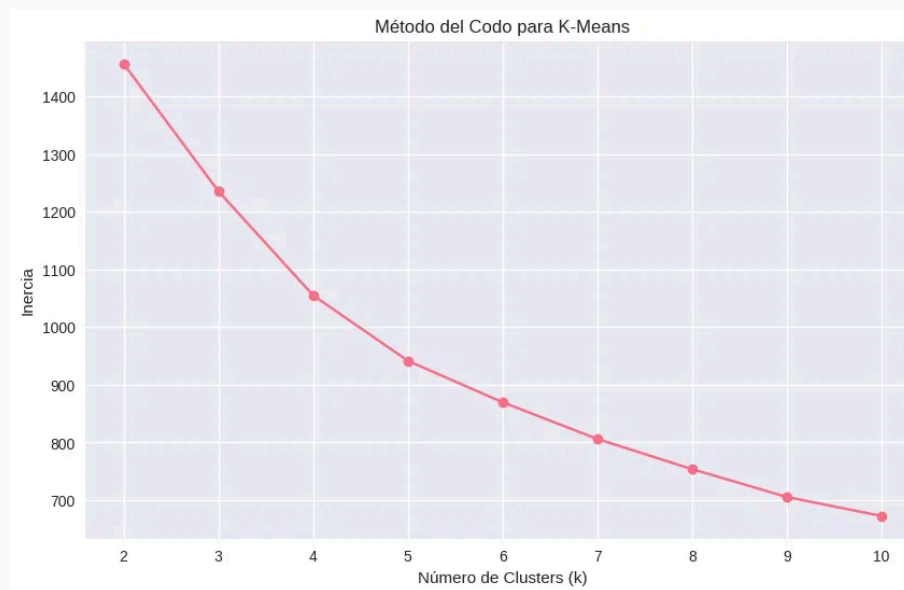
## Variables Analizadas:

 Calorías (media: 435.94)

 Carbohidratos (media: 35.07g)

 Azúcar (media: 9.05g)

 Proteínas (media: 24.15g)



# Determinación del Número Óptimo de Clusters



El **método del codo** nos muestra cómo disminuye la inercia (variabilidad dentro de los clusters) a medida que aumentamos el número de grupos.



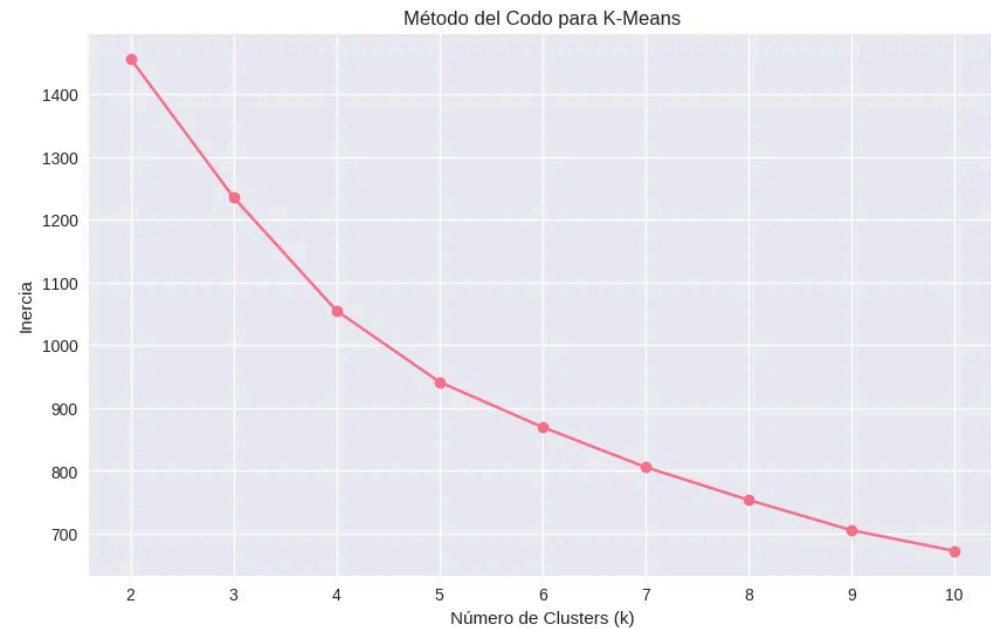
El **Silhouette Score** evalúa qué tan similares son las recetas dentro de su propio cluster comparado con otros clusters. Valores más altos indican mejor agrupación.



Basándonos en el Silhouette Score, determinamos que **2 clusters** es el número óptimo para nuestro conjunto de datos de recetas.



Un Silhouette Score de **0.214** indica una estructura de cluster razonable, aunque no fuertemente definida, lo que es común en datos nutricionales complejos.



# Resultados del Clustering



Identificamos **2 clusters** como número óptimo según el Silhouette Score (0.214).



El **Cluster 0** contiene 542 recetas (60.6%) y el **Cluster 1** contiene 353 recetas (39.4%).

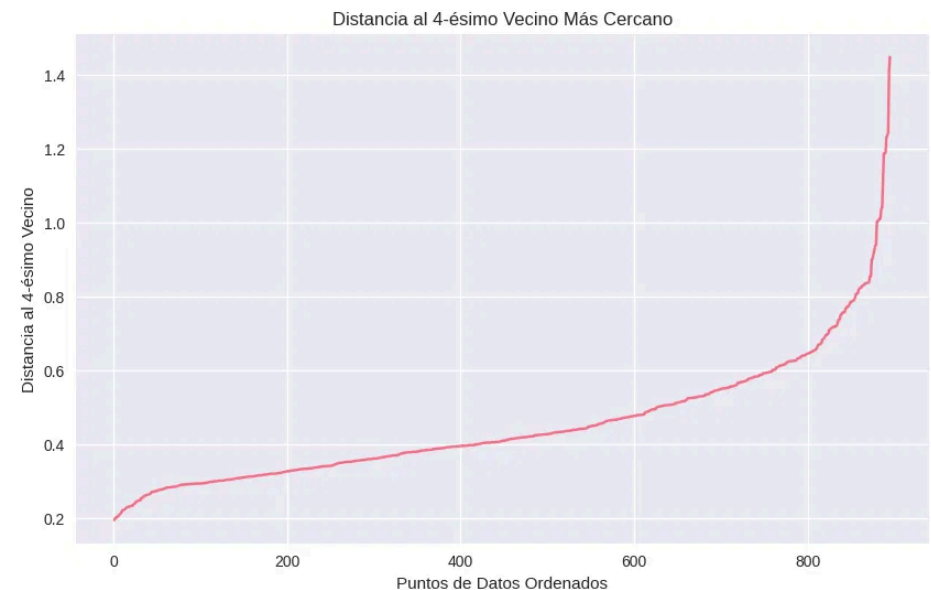
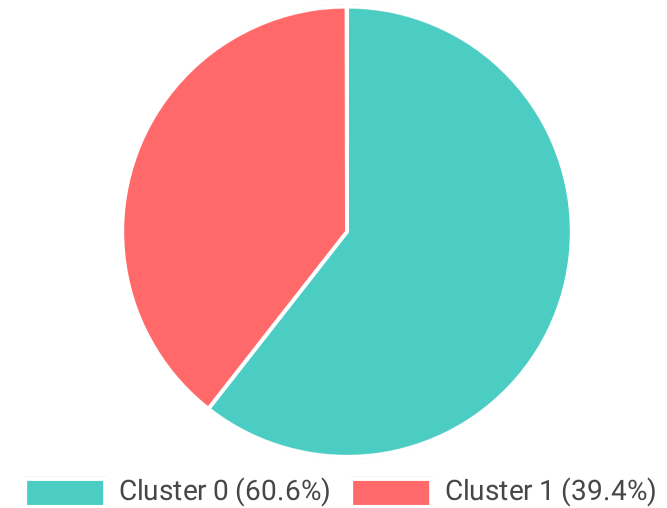


El algoritmo **K-Means** mostró el mejor rendimiento entre todos los algoritmos evaluados.

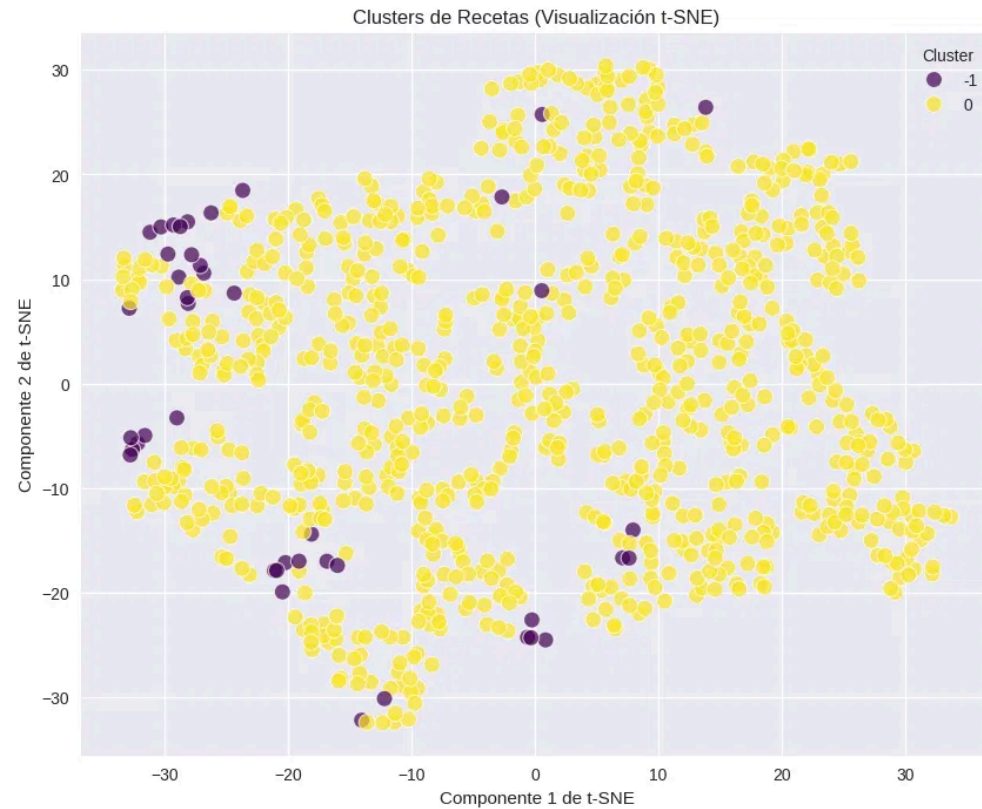


La visualización t-SNE muestra una clara separación entre los dos grupos de recetas identificados.

Distribución de Recetas por Cluster



# Características de los Clusters



## Cluster 0: Recetas Sustanciosas (60.6%)

🔥 Calorías:	629 kcal (Alto)
🍲 Carbohidratos:	36.5g
🍬 Azúcar:	5.5g (Bajo)
🥩 Proteínas:	34.5g (Alto)

*Platos principales y recetas saladas con alto contenido proteico, ideales para comidas completas.*

## Cluster 1: Recetas Ligeras (39.4%)

🔥 Calorías:	140 kcal (Bajo)
🍲 Carbohidratos:	32.9g
🍬 Azúcar:	14.5g (Alto)
🥩 Proteínas:	8.2g (Bajo)

*Postres, bebidas y snacks con alto contenido de azúcar, ideales para complementos o meriendas.*

# Conclusiones



## Dos perfiles nutricionales distintos

Nuestro análisis reveló dos grupos principales de recetas con características nutricionales claramente diferenciadas.



## Cluster 0: Recetas proteicas (60.6%)

Recetas con alto contenido calórico (629 cal) y proteico (34.5g), ideales para comidas principales y dietas altas en proteínas.



## Cluster 1: Recetas dulces (39.4%)

Recetas bajas en calorías (140 cal) y proteínas (8.2g), pero más altas en azúcar (14.5g), perfectas para postres y opciones ligeras.

## Aplicaciones Futuras



Desarrollo de sistemas de recomendación personalizados basados en perfiles nutricionales para sitios web de recetas.

