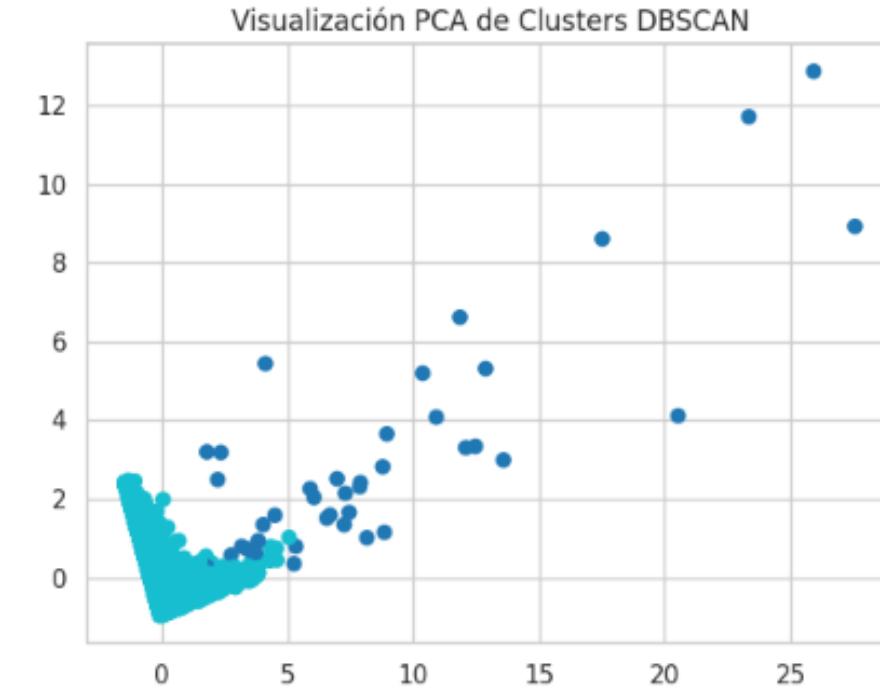
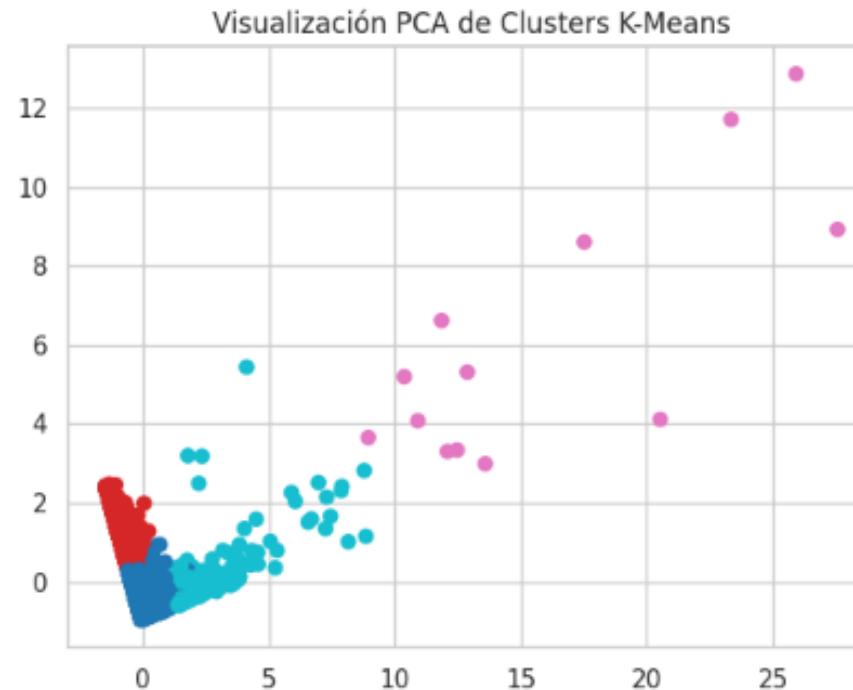


# ANÁLISIS COMPARATIVO ENTRE MODELOS

## ¿Qué diferencias clave surgieron entre los modelos?



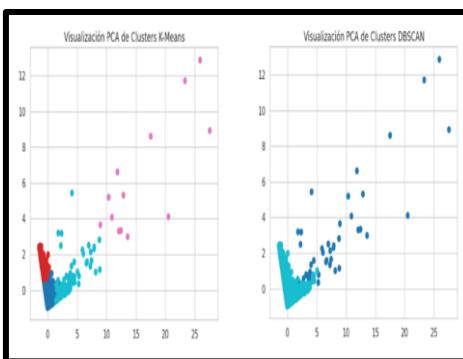
Al comparar ambas proyecciones PCA, se observa que K-means logra separar subgrupos de clientes, mientras que DBSCAN concentra la mayoría de los puntos en un único grupo, identificando principalmente valores atípicos.



**UEES**  
UNIVERSIDAD ESPÍRITU SANTO

## ANÁLISIS COMPARATIVO ENTRE MODELOS

### ¿Qué diferencias clave surgieron entre los modelos?



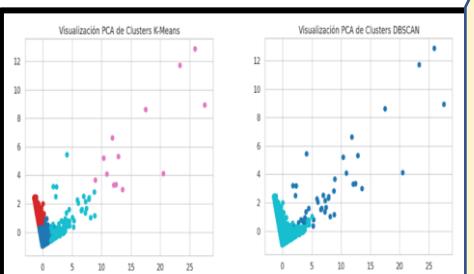
La comparación entre K-means y DBSCAN evidenció diferencias sustanciales en la forma de segmentar los datos.

K-means permitió dividir la base de clientes en cuatro clusters bien definidos, facilitando la identificación de perfiles diferenciados y jerarquizados según su valor RFM.

Este enfoque resultó adecuado para fines interpretativos y de segmentación estratégica.

## ¿Qué diferencias clave surgieron entre los modelos?

Por su parte, DBSCAN identificó un único grupo denso principal y clasificó a los clientes con comportamientos extremos como ruido (cluster -1).



Esto se refleja tanto en el número de clusters detectados como en las visualizaciones PCA y t-SNE, donde no se observa una separación clara de subgrupos internos.

En consecuencia, DBSCAN se mostró más efectivo para la detección de atípicos que para la segmentación detallada de clientes en este conjunto de datos.



### ¿Qué tipo de perfiles se pueden identificar?



A partir del análisis RFM y la segmentación mediante K-means ( $k = 4$ ), se identificaron cuatro perfiles de clientes claramente diferenciados:

Un primer grupo corresponde a clientes inactivos o de bajo valor, caracterizados por una recencia muy elevada (alrededor de 250 días), una frecuencia de compra mínima y un gasto monetario bajo.

Un segundo perfil agrupa clientes ocasionales, con recencia intermedia ( $\approx 40$  días), baja frecuencia y un gasto moderado.

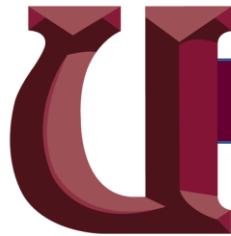


### ¿Qué tipo de perfiles se pueden identificar?



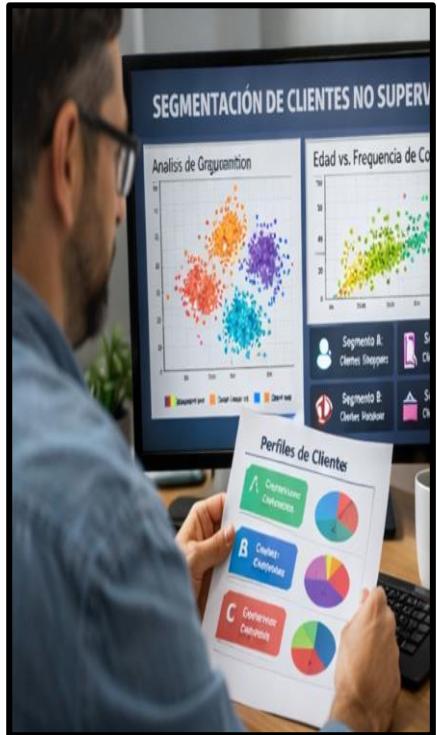
Asimismo, se identificó un grupo de clientes frecuentes y valiosos, con recencia baja ( $\approx 15$  días), alta frecuencia de compra (más de 20 transacciones) y un gasto monetario considerable.

Finalmente, se observa un grupo reducido de clientes de muy alto valor, con recencia mínima (menos de 10 días), una frecuencia extremadamente alta (más de 80 compras) y un valor monetario muy superior al promedio general, representando el segmento más relevante desde el punto de vista económico.



## CONCLUSIONES Y RECOMENDACIONES

### ¿Qué limitaciones se encontraron y cómo podrían abordarse?



Entre las principales limitaciones se identificó la sensibilidad de DBSCAN a la selección de parámetros, particularmente el valor de  $\text{eps}$ , lo que condiciona directamente la cantidad de clusters detectados.

Asimismo, al eliminar los valores atípicos, el modelo conserva únicamente un perfil promedio, perdiendo capacidad para distinguir subsegmentos relevantes.

En el caso de K-means, una limitación importante es la asignación forzada de todos los clientes a un cluster, incluso cuando presentan comportamientos extremos.



### ¿Qué limitaciones se encontraron y cómo podrían abordarse?



Como líneas de mejora, se podría explorar el uso de algoritmos alternativos como HDBSCAN o clustering jerárquico, así como evaluar transformaciones de las variables RFM para reducir la influencia de valores extremos.

Además, un análisis RFM en distintos períodos temporales permitiría incorporar una dimensión evolutiva del comportamiento del cliente.