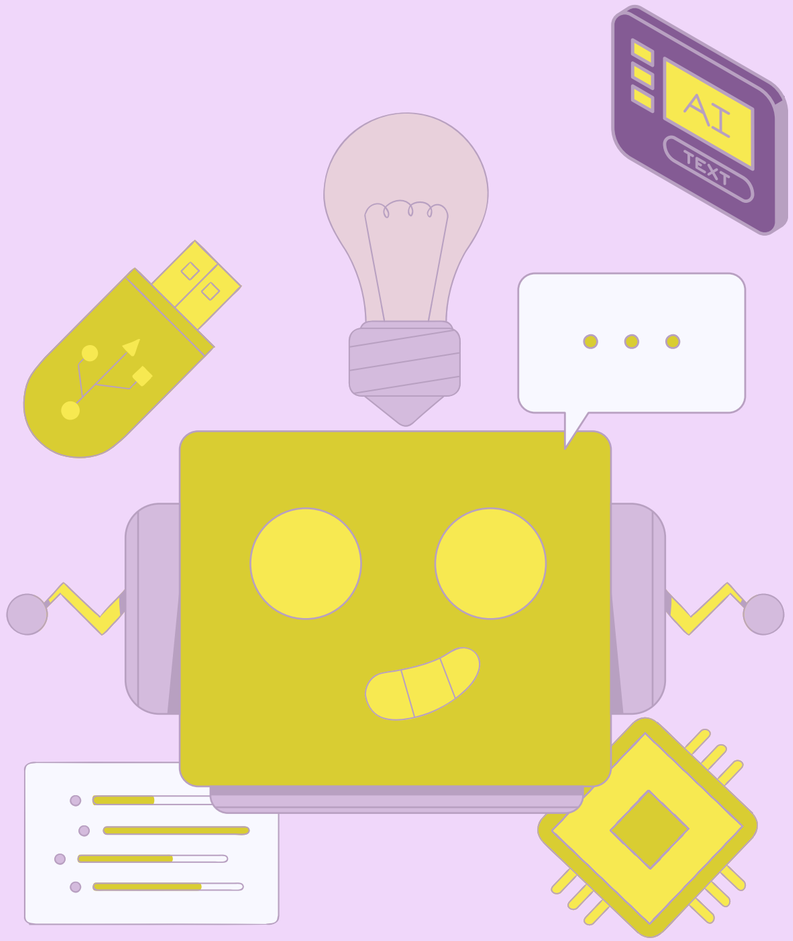
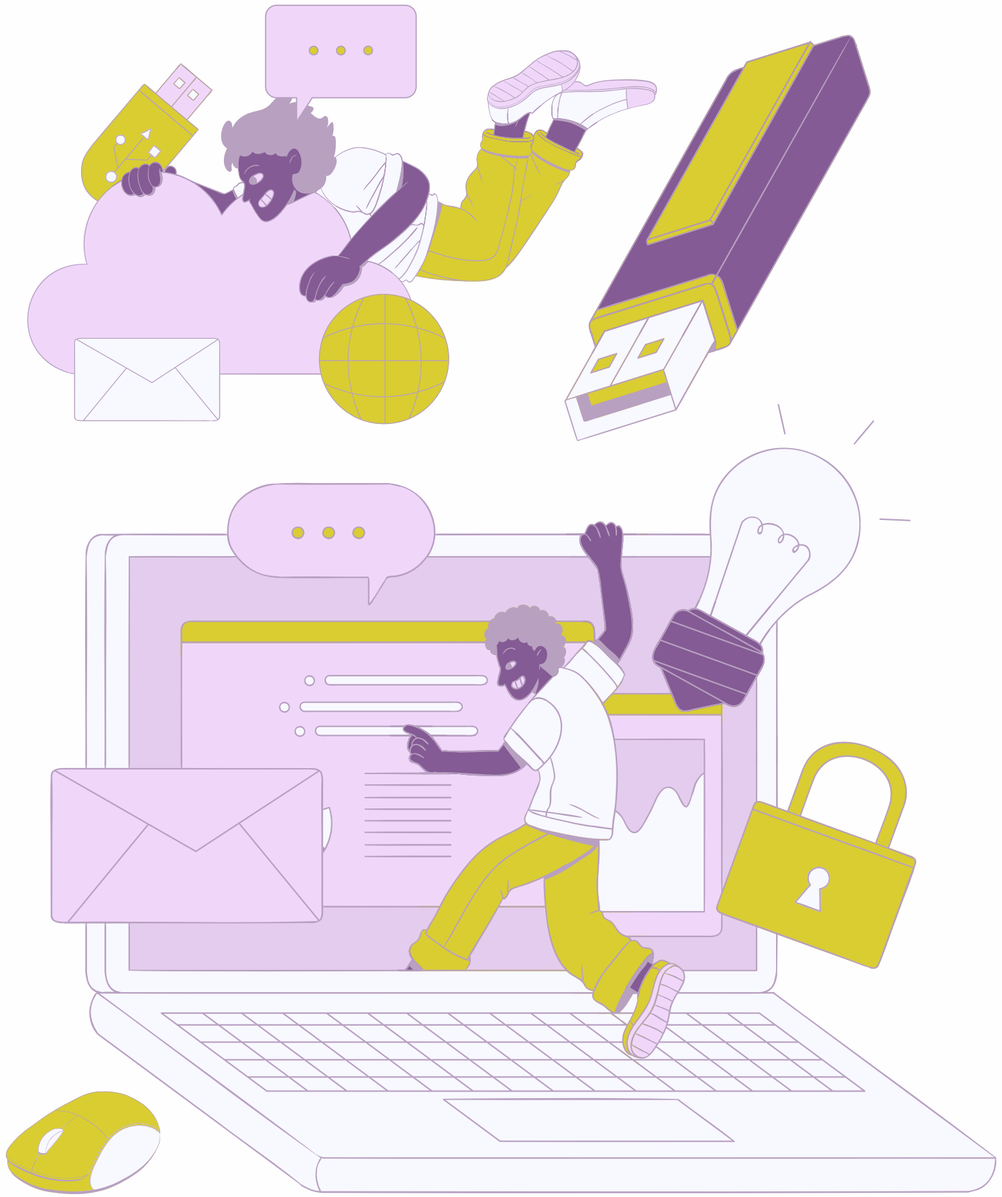



# SEGMENTACIÓN DE CLIENTES CON MACHINE LEARNING

Patricia Díez

Abril 2025



# ÍNDICE

# INTRODUCCIÓN

## OBJETIVO DEL PROYECTO

La segmentación de los clientes mayoristas, basado en sus patrones de compras haciendo uso de Machine Learning para ayudarnos.

## ¿ PARA QUÉ SIRVE?

Lo que esta segmentación permitirá será llevar a cabo distintas estrategias de marketing para cada grupo creado.

# DESCRIPCIÓN DEL DATASET

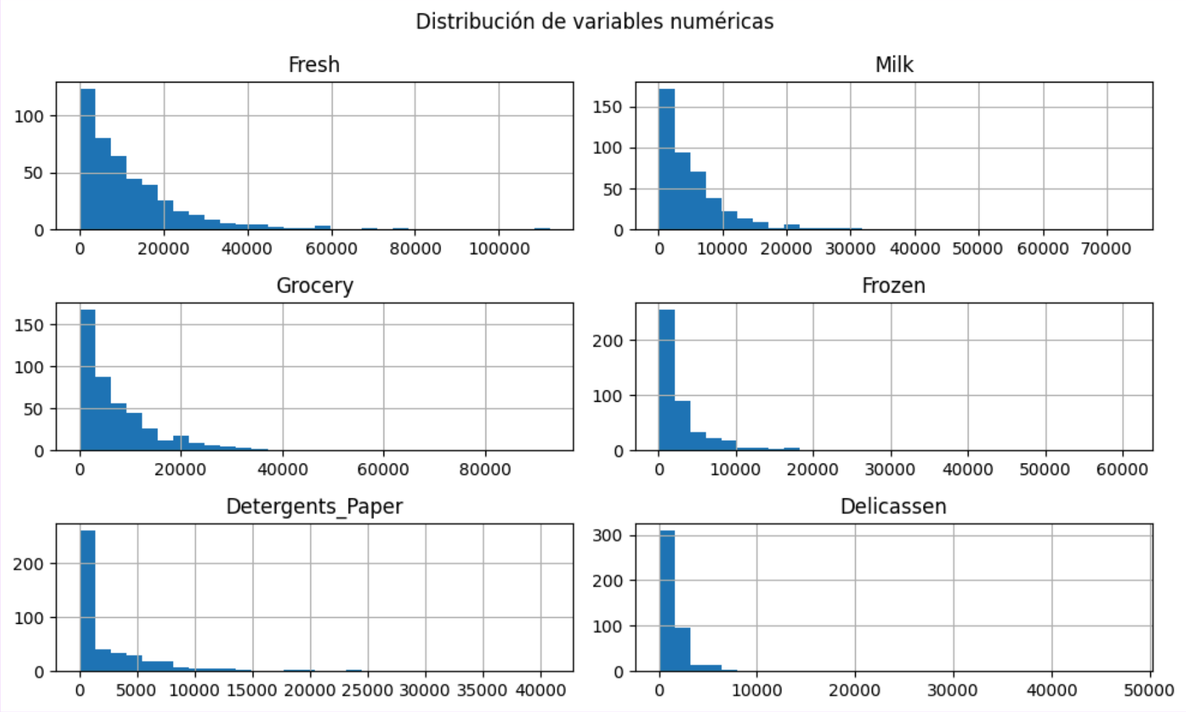
## WHOLESALE CUSTOMERS

- Dataset público de Kaggle

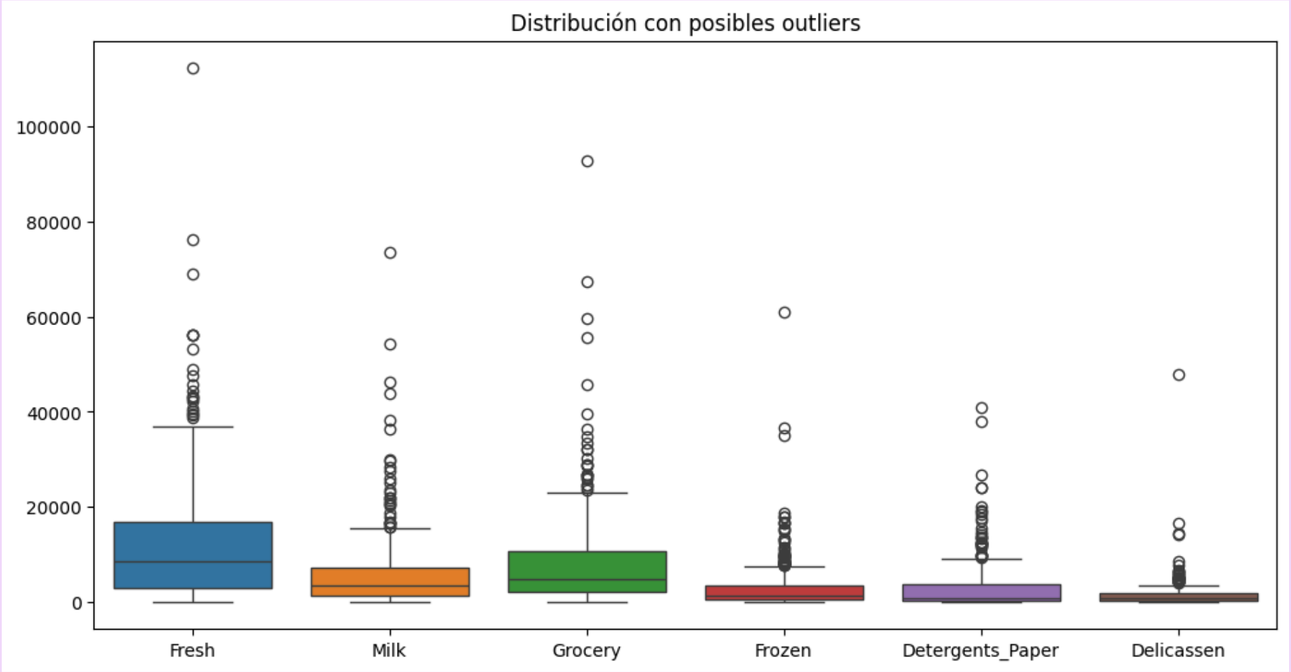
- Contiene información sobre el gasto por 8 categorías de 440 clientes mayoristas.

- Variables recogidas:
  - Variables numéricas: Fresh, Milk, Grocery, Frozen, Detergents_Paper, Delicassen
  - Variables categóricas: Region y Channel

# PREPROCESAMIENTO Y EDA

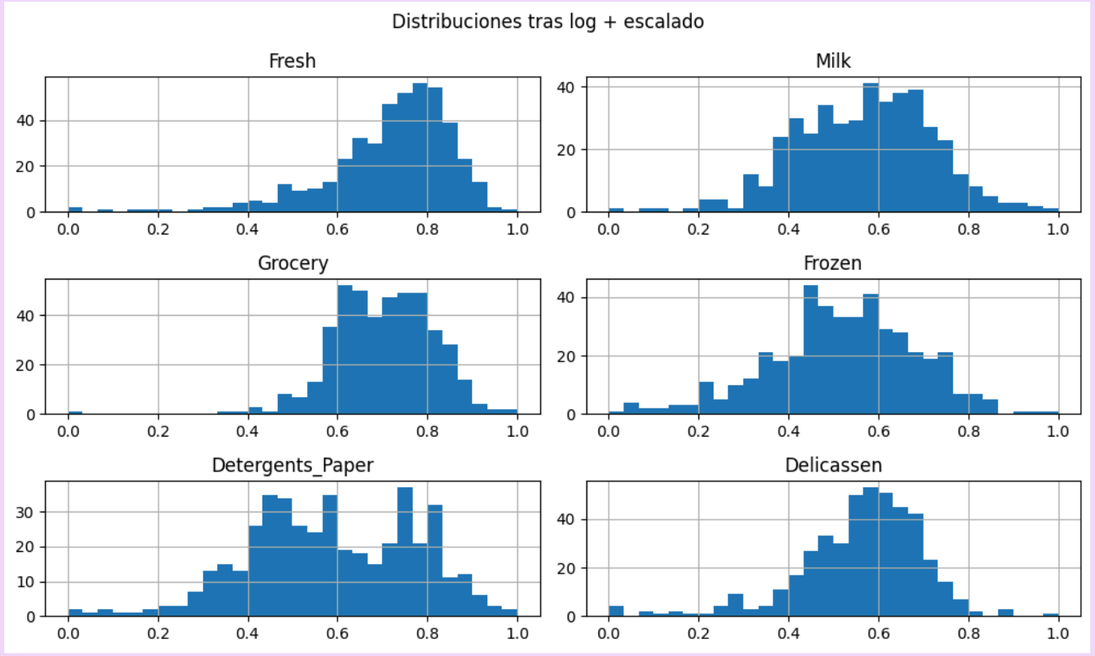



## MUCHO SESGO HACIA LA DERECHA

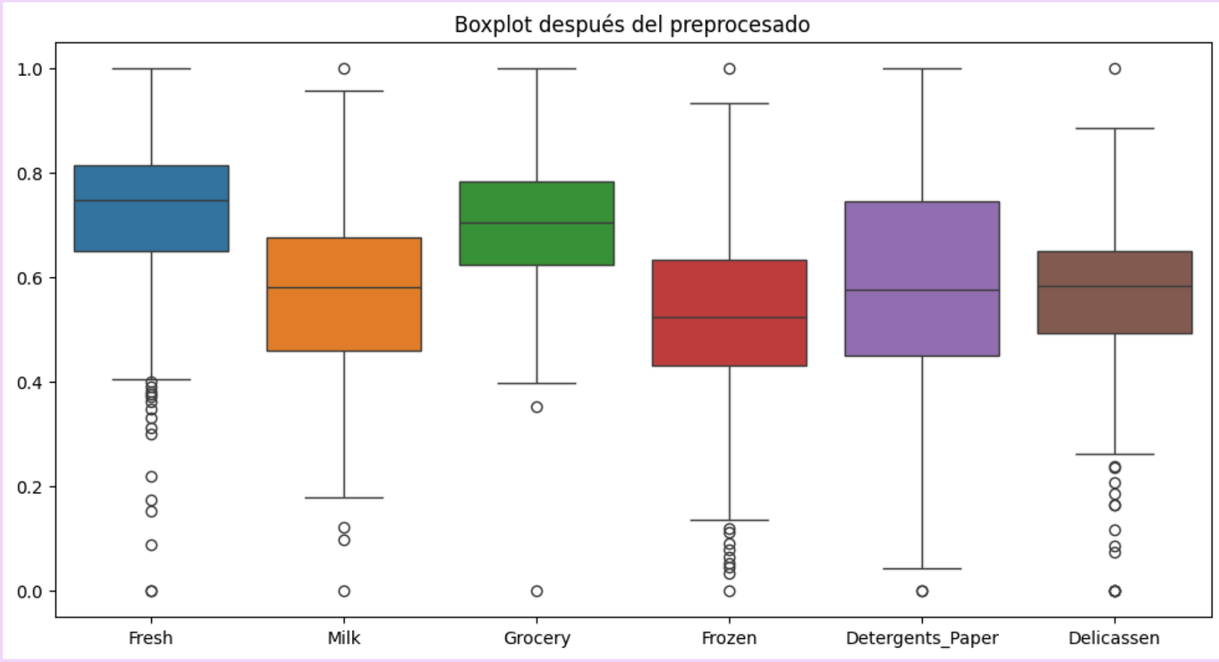



## MUCHOS OUTLIERS

# PREPROCESAMIENTO Y EDA

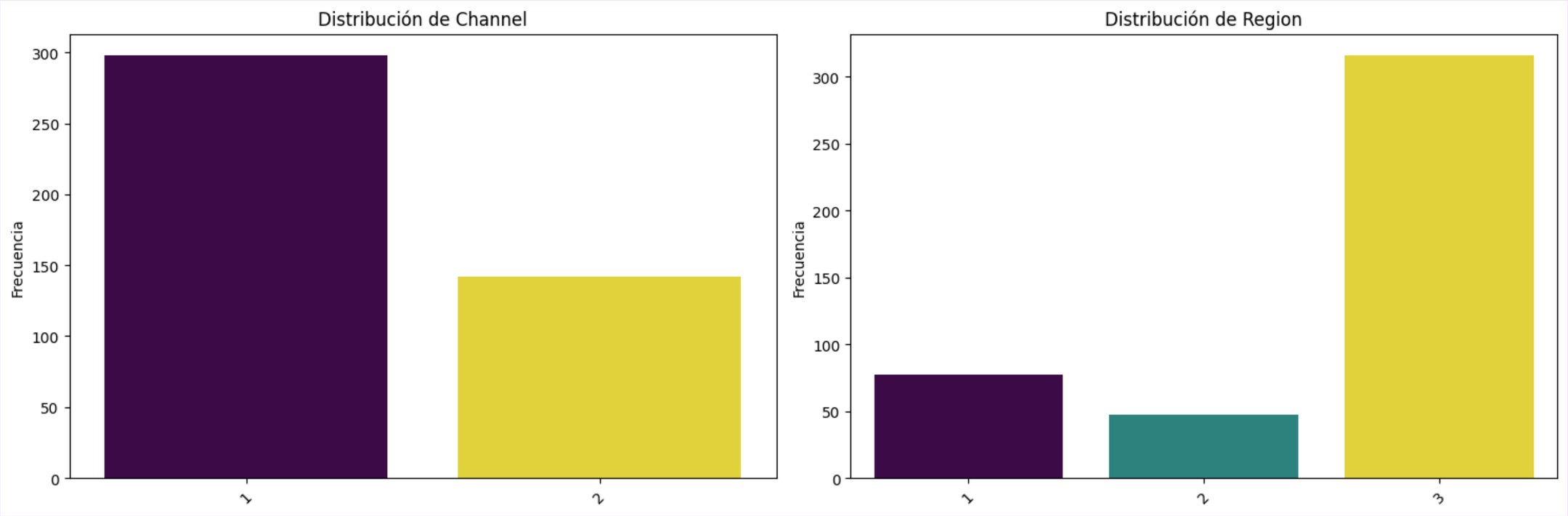



## LOGARITMO




## ESCALADO

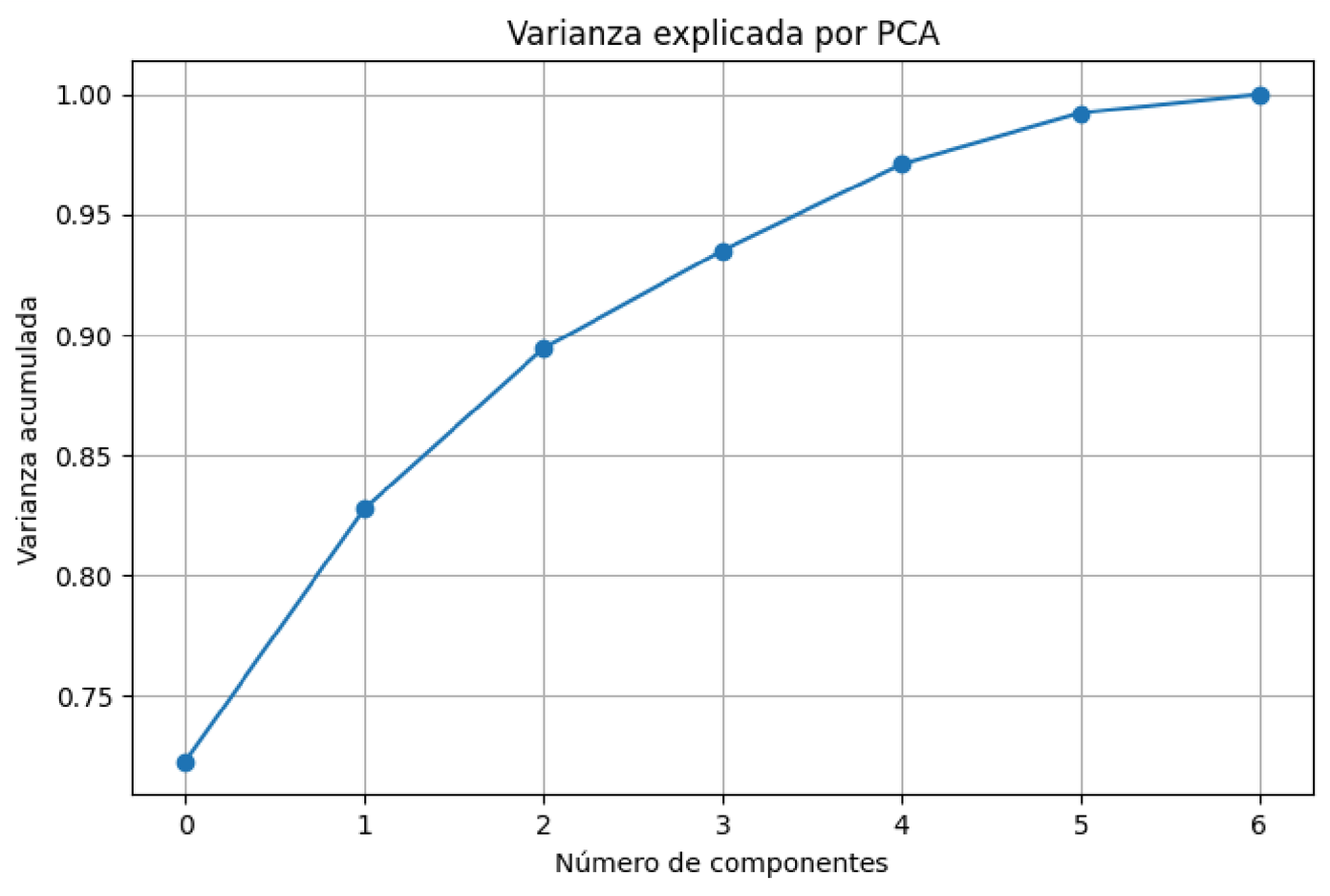
# PREPROCESAMIENTO Y EDA




ONE - HOT ENCODING DE CHANNEL

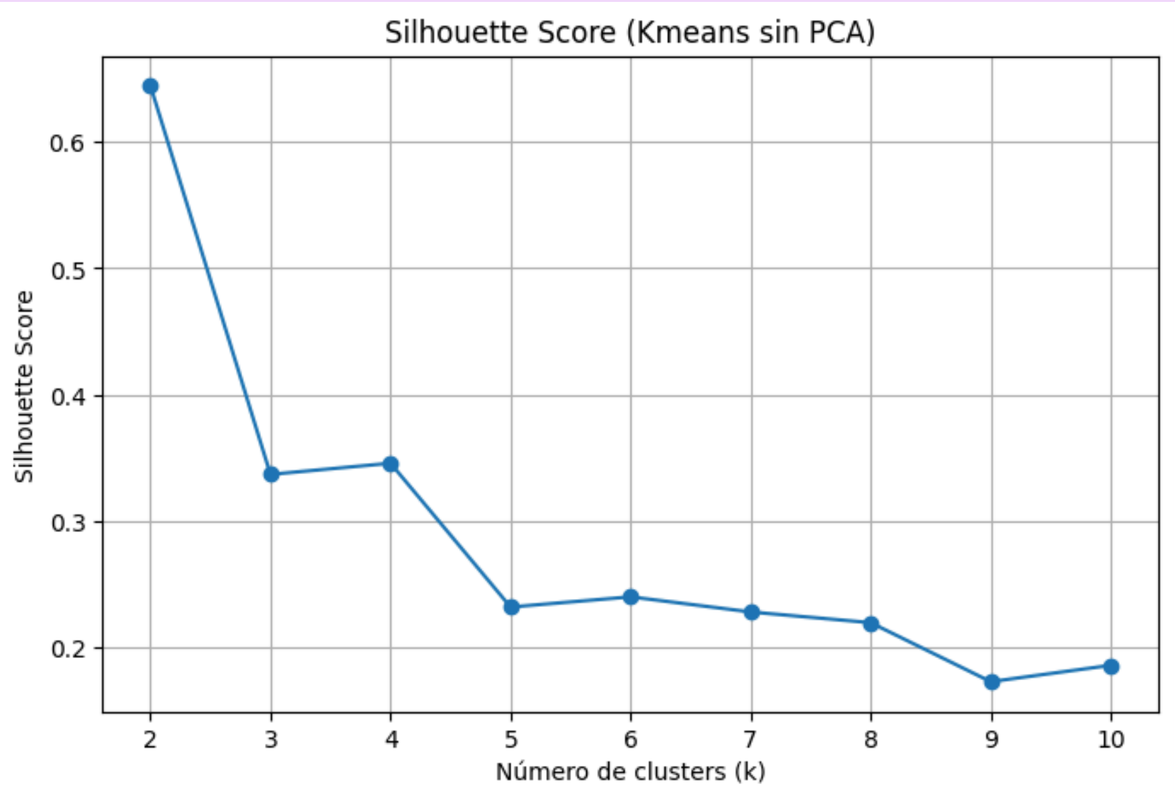
ELIMINAMOS REGION

# REDUCCIÓN DE DIMENSIONALIDAD (PCA)




- Con 2 componentes: se explica el 89% de la varianza total
- Con 3 componentes: el 94%
- Con 4 componentes: el 97%

# K = 2 EN KMEANS

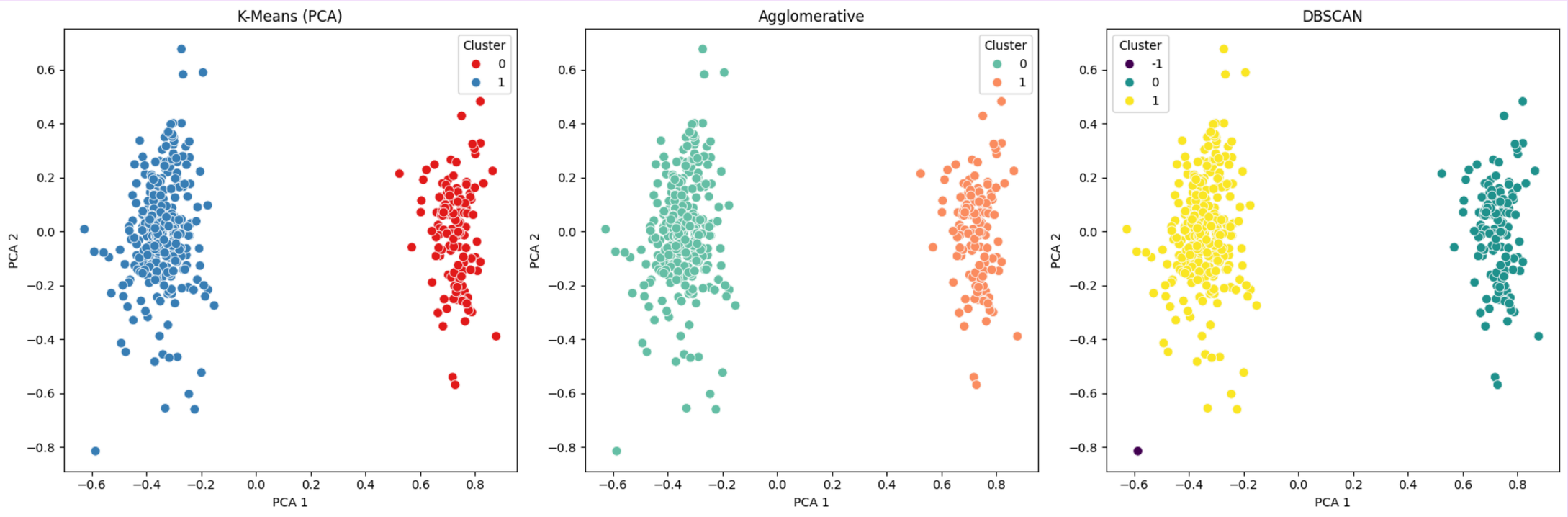



- Probamos con valores de k entre 2 y 10
- Método del codo: no ha sido concluyente
- Silhouette Score: el valor máximo con k=2

# COMPARACIÓN DE RESULTADOS

| Algoritmo | Silhouette Score |
|---|---|
| KMeans (sin PCA) | 0.64 |
| KMeans (con PCA) | 0.69 |
| DBSCAN | 0.79 |
| Agglomerative | 0.78 |

# ELECCIÓN DEL MODELO FINAL

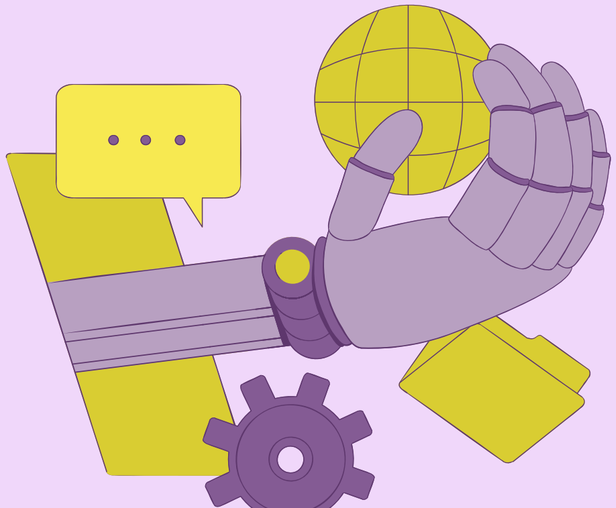



## MODELO FINAL ELEGIDO = KMEANS CON PCA

# CONCLUSIONES

- Probamos 3 modelos: KMeans, DBSCAN y Agglomerative

- Aunque DBSCAN  y Agglomerative han tenido mejor Silhouette score, elegimos KMeans con PCA como modelo final:
  - Simplicidad y velocidad: Rápido de entrenar y fácil de interpretar.
  - Estabilidad: No depende de parámetros sensibles como en DBSCAN.
  - Generalización: Es aplicable a nuevos datos
  - Interpretabilidad: Permite analizar bien las variables de cada grupo



# RESULTADOS

**CLUSTER 0**

GROCERY

DELICASSEN

DETERGENTS

MILK

→ clientes tipo Retail

FRESH

→ perfil Horeca

FROZEN

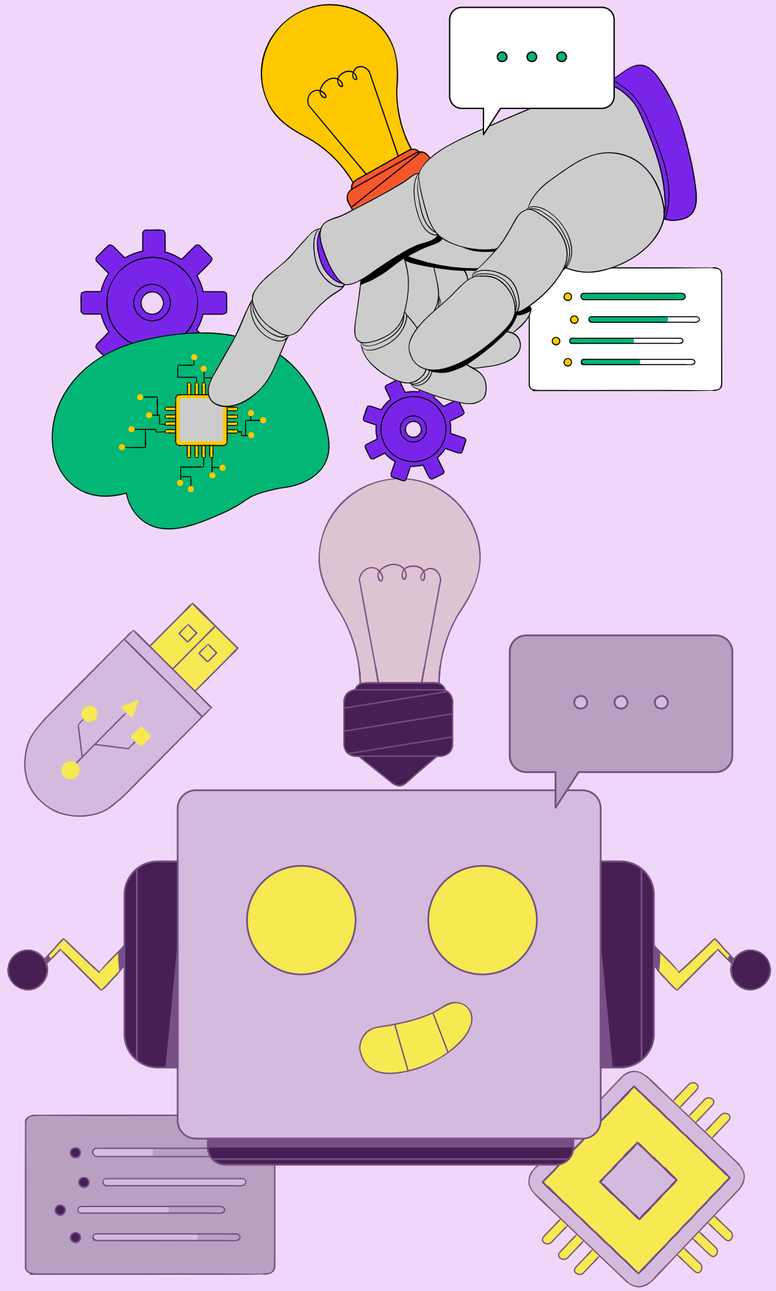
**CLUSTER 1**

# POSIBLES MEJORAS

Este sistema es útil para segmentación y estrategia de marketing → futuras mejoras incluirían:

OPTIMIZACIÓN DE PARÁMETROS

TRATAMIENTO DE OUTLIERS



# GRACIAS!