

# Clasificación de Vino con PCA y KMeans

Jeiner Cantillo, Robert Lopez, Santiago Carmona – Vinos Zungo S.A.

IES INFOTEP - Abril 2025

## Introducción

La empresa **Vinos Zungo S.A.** ha lanzado un nuevo vino elaborado a base de banano. Tras un análisis de laboratorio, se identificaron sus propiedades químicas, y se busca determinar a cuál de los tres tipos de vino conocidos se asemeja más, utilizando técnicas de aprendizaje no supervisado como el Análisis de Componentes Principales (PCA) y el algoritmo KMeans.

## Objetivos

- Implementar un modelo PCA desde cero para reducir la dimensionalidad del dataset.
- Visualizar los datos en un espacio de menor dimensión para facilitar su análisis.
- Aplicar KMeans para clasificar los vinos en grupos.
- Determinar a qué grupo pertenece el vino de banano.
- Evaluar cómo influye la selección de componentes principales en la clasificación.

## Desarrollo del Trabajo

Se utilizó el dataset clásico de vinos que contiene 178 muestras de tres clases distintas, con 13 características químicas. Se desarrolló una clase personalizada en Python para implementar PCA desde cero, calculando los valores y vectores propios de la matriz de covarianza de los datos estandarizados.

Posteriormente, se proyectaron los datos en los 2 y 3 componentes principales más relevantes. Se entrenó el modelo KMeans sobre esta proyección y se clasificó el nuevo vino de banano comparando su ubicación en el espacio reducido con los centroides de los clusters.

# Resultados

## Imagen 1 – Visualización PCA con 2 Componentes

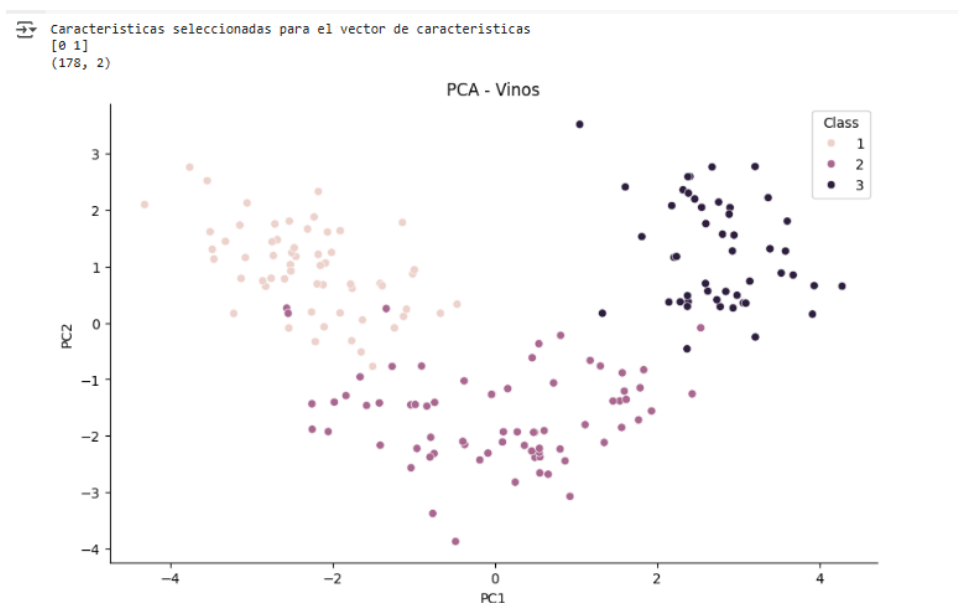


Figura 1: Proyección PCA de los vinos originales usando 2 componentes.

Esta visualización muestra una clara separación entre las tres clases de vinos cuando se utilizan los dos componentes principales más relevantes.

## Imagen 2 – Clasificación del Vino de Banano

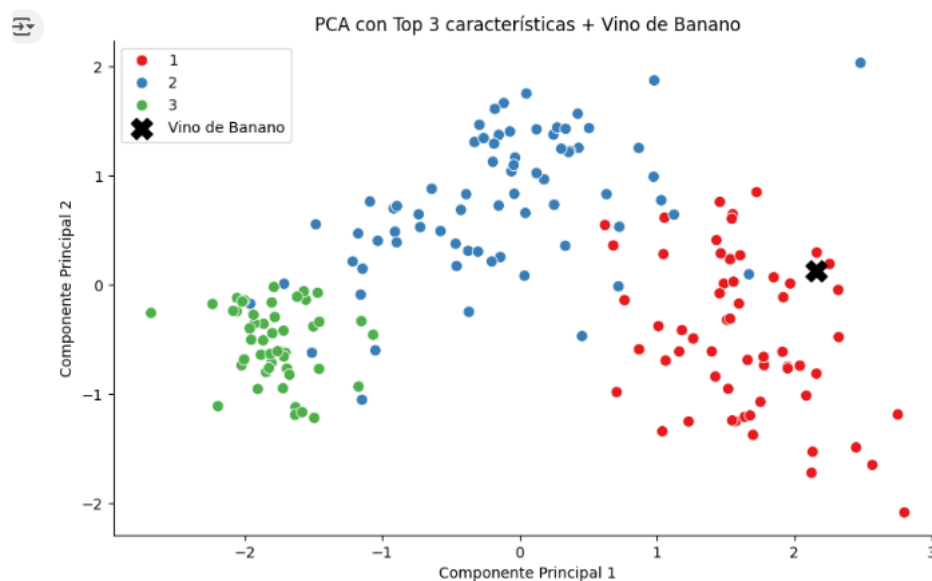


Figura 2: Clusters generados con KMeans usando 3 componentes. El vino de banano aparece con una X negra.

Se observa que el nuevo vino de banano cae dentro del grupo identificado como clase 1 (rojo), lo que indica una fuerte similitud química con ese tipo de vino.

## Conclusiones

- PCA permitió reducir la dimensionalidad del problema, facilitando la visualización y clasificación.
- KMeans fue eficaz para agrupar los vinos en función de sus propiedades.
- El vino de banano fue clasificado dentro del grupo de clase 1, lo que sugiere similitudes químicas significativas.
- Cambiar la cantidad o calidad de los componentes usados afecta la clasificación, mostrando la importancia de una buena selección de variables.

## Referencias

- Scikit-learn documentation: <https://scikit-learn.org/>
- Matplotlib documentation: <https://matplotlib.org/>
- Dataset de vinos: UCI Machine Learning Repository