

INFORME DE ANÁLISIS PCA + K-MEANS

Fecha de Generación: 2025-10-21 19:09:47

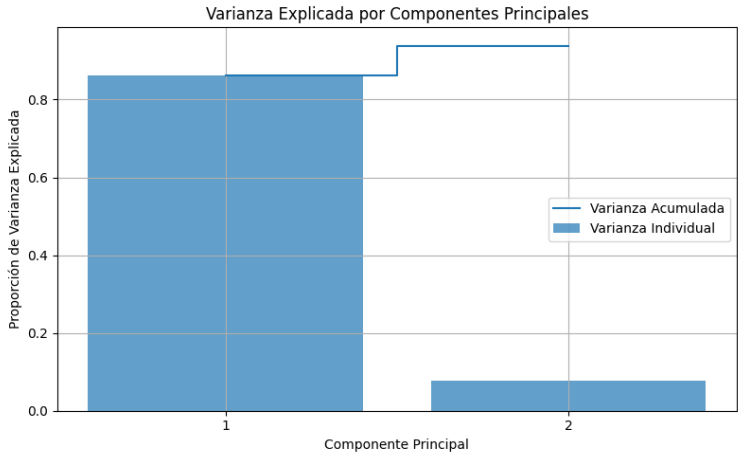
1. ANÁLISIS DE COMPONENTES PRINCIPALES (PCA)

El modelo PCA redujo la dimensionalidad a 2 Componentes Principales (PC1, PC2), capturando un total de 93.87% de la varianza original en los datos químicos y de efectos. Esto asegura que la visualización 2D es altamente representativa de la estructura real de los datos.

Varianza Individual por Componente:
- PC1: 86.15% - PC2: 7.71%

Conclusión: Interpretación de PC1 (Eje de Máxima Variación)
PC1 está dominada por: 'Uplifted' (Correlación POSITIVA); 'Limonene' (Correlación POSITIVA); 'Sleepy' (Correlación NEGATIVA); . Este eje define la principal diferencia entre las cepas.

Gráfico de Varianza Explicada (Ver archivo: pca_variance_plot.png):



INFORME DE ANÁLISIS PCA + K-MEANS

2. RESULTADOS DEL CLUSTERING K-MEANS (K=3)

Perfiles de Características Promedio por Clúster:

Cluster	THC_Level	Myrcene	Limonene	Caryophyllene	Relaxed	Happy	Uplifted	Sleepy
0.0	21.0	0.25	0.35	0.28	4.5	3.5	3.0	2.5
1.0	18.0	1.07	0.08	0.62	5.0	2.67	1.0	4.33
2.0	24.0	0.25	0.78	0.18	2.5	5.0	5.0	1.0

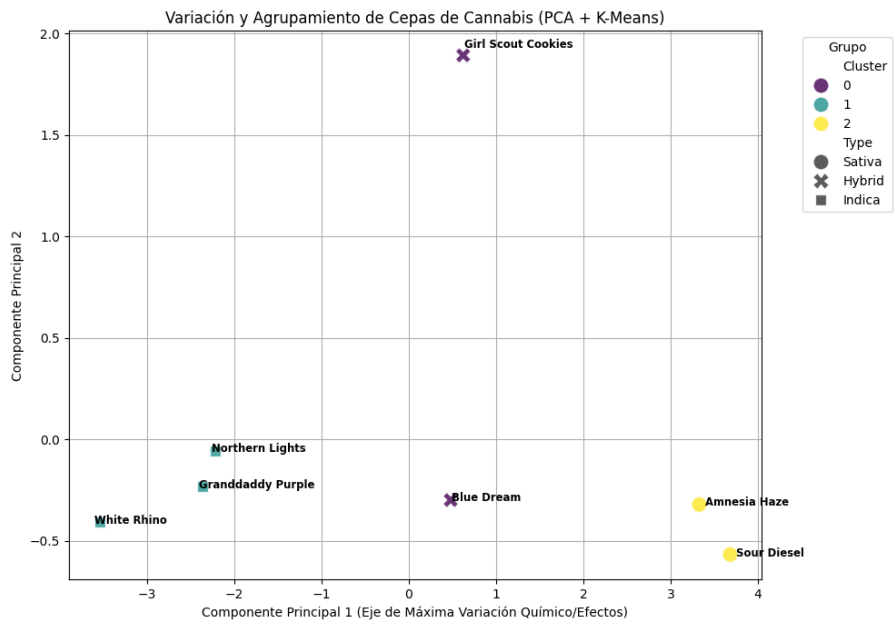
Interpretación de los Clústeres (Basado en Perfiles):

- Clúster 0 (2 Cepas): Identificado como ****SATIVA-DOMINANTE****. Característica clave: Alto Limonene y Alto efecto Energético (3.00).
- Clúster 1 (3 Cepas): Identificado como ****INDICA-DOMINANTE****. Característica clave: Alto Myrcene (1.07) y Alto efecto Sedante (4.33).
- Clúster 2 (2 Cepas): Identificado como ****SATIVA-DOMINANTE****. Característica clave: Alto Limonene y Alto efecto Energético (5.00).

3. VALIDACIÓN Y VISUALIZACIÓN

La validación visual es crucial. El gráfico de dispersión PCA/K-Means (f7) muestra cómo los 3 clústeres forman grupos distintos en el espacio 2D, confirmando que el clustering ha encontrado la variación estructural basada en la composición química.

Gráfico de Agrupamiento PCA/K-Means (Ver archivo: `pca_kmeans_clustering_plot.png`):



CONCLUSIÓN FINAL Y RECOMENDACIÓN

El análisis demostró que la principal variación entre las cepas se relaciona directamente con el contenido de Terpenos (ej. Myrcene/Limonene) y sus efectos percibidos (Sedante/Energético). El modelo K-Means ha segmentado exitosamente el dataset en 3 grupos que reflejan esta variación biológica, lo cual es útil para la recomendación de productos y el desarrollo de cepas con perfiles de efectos específicos.