**Descubriendo Perfiles Ocultos en Datos de Vinos**

**Introducción**

Objetivo de la evaluacion: Explorar un conjunto de datos de análisis químico de vinos para descubrir patrones, perfiles y correlaciones entre diferentes componentes químicos.

* Conjunto de Datos: [Enlace al conjunto de datos](https://storage.googleapis.com/the_public_bucket/wine-clustering.csv).
* Exploración de Datos
* Distribución y Relación de Atributos
* Análisis de Distribución:
* Visualización de la distribución de atributos clave como Alcohol, Ácido Málico, Flavonoides, etc.
* Uso de herramientas visuales y estadísticas para comprender la variabilidad en los datos.
* Clustering Analysis
* Técnicas de Agrupamiento:
* Aplicación de K-Means, Agrupamiento Jerárquico, DBSCAN.
* Determinación del número óptimo de clústeres.

Descripcion del Trabajo realizado:

**1. Obtención de Datos desde una API de Google:**

Iniciamos nuestro proyecto accediendo a una API de Google para obtener datos sobre vinos, incluyendo información química.

**2. Transformación de Datos:**

Convertimos los datos obtenidos a un DataFrame utilizando la biblioteca pandas.

Exploramos las primeras filas del DataFrame para comprender la estructura de los datos.

**3. Preparación de Datos:**

Revisamos los tipos de datos para garantizar la consistencia.

Abordamos cualquier valor nulo o atípico que podría afectar el análisis, eliminándolos para mantener la integridad de los resultados.

**4. Análisis de Distribución y Correlación:**

Utilizamos herramientas visuales y estadísticas para explorar cómo se distribuyen y correlacionan las variables.

Calculamos una matriz de correlación para comprender las relaciones entre diferentes variables.

**5. Escalado de Datos y Reducción Dimensional (PCA):**

Escalamos los datos para evitar sesgos debidos a diferentes escalas.

Aplicamos el análisis de componentes principales (PCA) para reducir la dimensionalidad y conservar la información relevante.

**6. Evaluación de Algoritmos de Clustering:**

Implementamos algoritmos de clustering como K-Means y Agglomerative Clustering.

Tras evaluaciones, determinamos que Agglomerative Clustering ofrece resultados óptimos para agrupar los datos.

**7. Evaluación de Variables Relevantes:**

Identificamos y evaluamos un par de variables clave (Alcohol y Flavonoides) que contribuyen significativamente a la formación de clústeres.

Visualizamos la distribución de clústeres utilizando las variables seleccionadas.

Resumen final del trabajo:

**1. Clusters y Tipos de Vino:**

Al aplicar técnicas de clustering, identificamos tres clusters distintos en los datos de vinos.

Sorprendentemente, estos clusters se corresponden con las categorías de vinos blanco, rosado y tinto.

**2. Correlación Inesperada: 'OD280' y 'Alcohol':**

Descubrimos una correlación inesperada entre dos variables aparentemente diferentes: 'OD280' (absorbancia de proteínas) y 'Alcohol'.

Los vinos con mayor contenido de alcohol también tienden a mostrar una absorbancia específica de proteínas ('OD280'). Esto podría indicar una relación compleja entre el contenido de alcohol y ciertos compuestos proteicos.

**3. Correlación Inesperada: 'Proline' y 'OD280':**

Otro hallazgo intrigante es la correlación positiva entre la 'Proline' (un aminoácido) y la absorbancia de proteínas ('OD280').

Este patrón sugiere que la concentración de Proline podría influir en la absorbancia de proteínas, lo cual podría tener implicaciones para la calidad del vino.

**4. Correlación Inesperada: 'Color\_Intensity' y 'Flavanoids':**

Observamos una correlación entre la intensidad de color y la cantidad de flavonoides en los vinos.

Los vinos con mayor intensidad de color tienden a tener una mayor presencia de flavonoides, lo cual puede ser un factor importante en la percepción sensorial y la calidad del vino.

**Implicaciones Prácticas:**

**Producción de Vino:**

Los productores de vino pueden utilizar esta información para ajustar las proporciones de compuestos químicos en sus procesos de producción, optimizando así la calidad del producto final.

**Marketing y Preferencias del Consumidor:**

La identificación de perfiles de vino distintos puede influir en estrategias de marketing y en la satisfacción de los consumidores. Los entusiastas del vino podrían encontrar útiles estas categorías al seleccionar vinos según sus preferencias.

**Investigación Adicional:**

Los hallazgos inesperados abren la puerta a investigaciones adicionales sobre las interacciones complejas entre los componentes químicos del vino. Esto podría conducir a avances en la comprensión de los factores que contribuyen a la calidad y el sabor del vino.