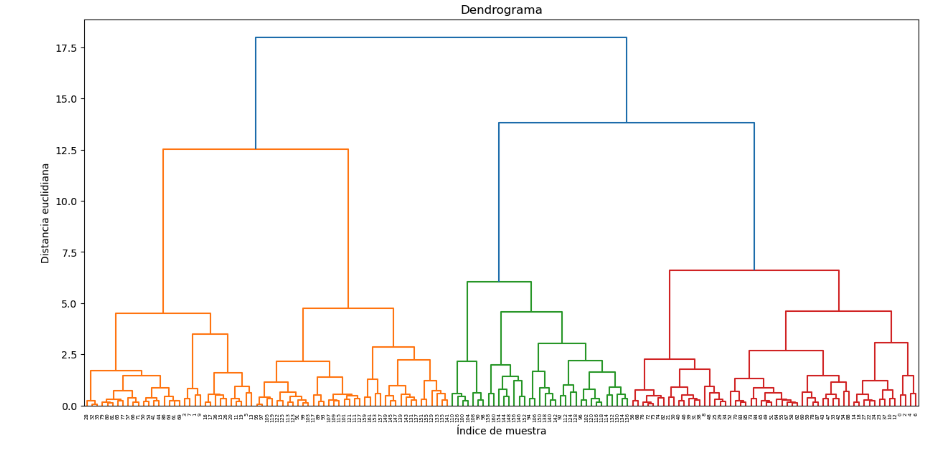
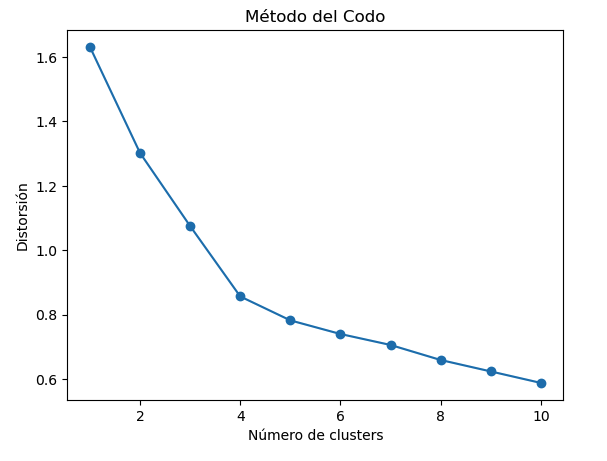
**Interpretación de resultados**

**Modelo\_Hierarchical\_Clustering**

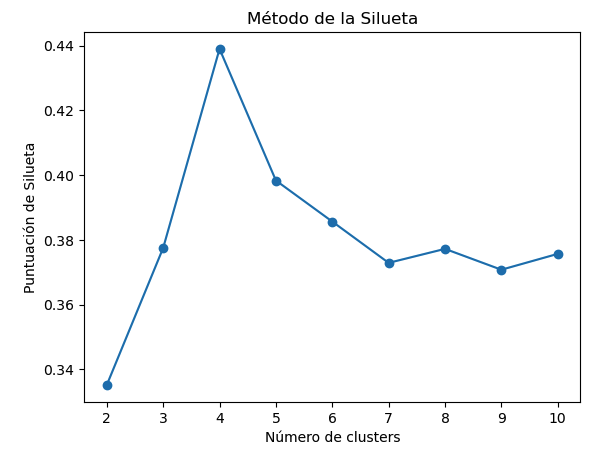


El dendograma nos da una idea de cómo están estructurados los datos y cuantos grupos naturales pueden existir en ellos. Cada ramificación representa un grupo de datos y cada unión indica la fusión de dos grupos en uno solo. En este caso tenemos 1 grupo principal que se divide en dos y de ahí salen otros subgrupos adicionales, En el caso de las fusiones que ocurren en alturas más altas como es el caso de la azul y la naranja, los datos tienden a ser menos similares.

**Identificación de números óptimos de clusters**

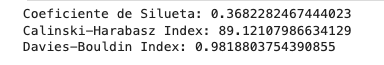


Mediante el método del codo se determinan el número óptimo de clústeres en los datos mediante el algoritmo K-Means, en el punto donde la distorsión comienza a disminuir más lentamente, lo que se asemeja a un "codo". Este punto sugiere el número óptimo de clusters. En este caso el punto ideal o número de clústers óptimos sería el 4.



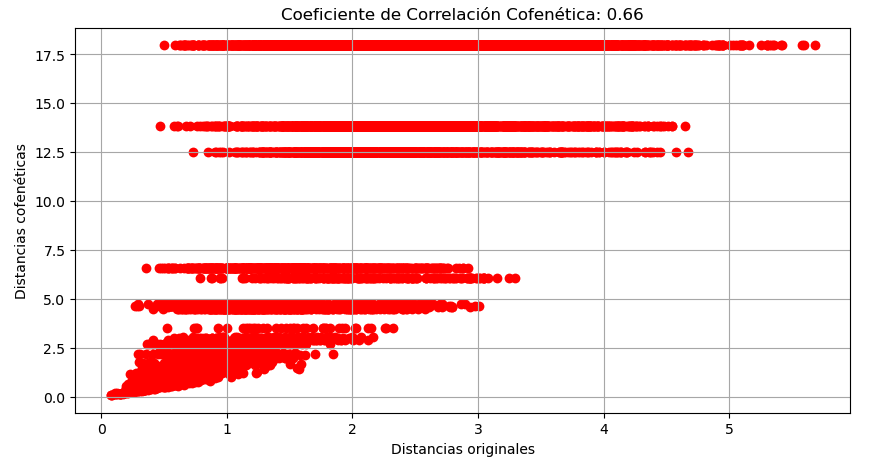
Mediante el método de la puntuación de silueta, también se determina el número óptimo de clústeres en los datos mediante el algoritmo K-Means. El grafico muestra el número de clusters en el eje x y la puntuación de silueta en el eje y. dada que la puntuación más alta la tiene el número de clusters 4 se confirma por otro método que esta es el número de clusters óptimos.

**Metricas**

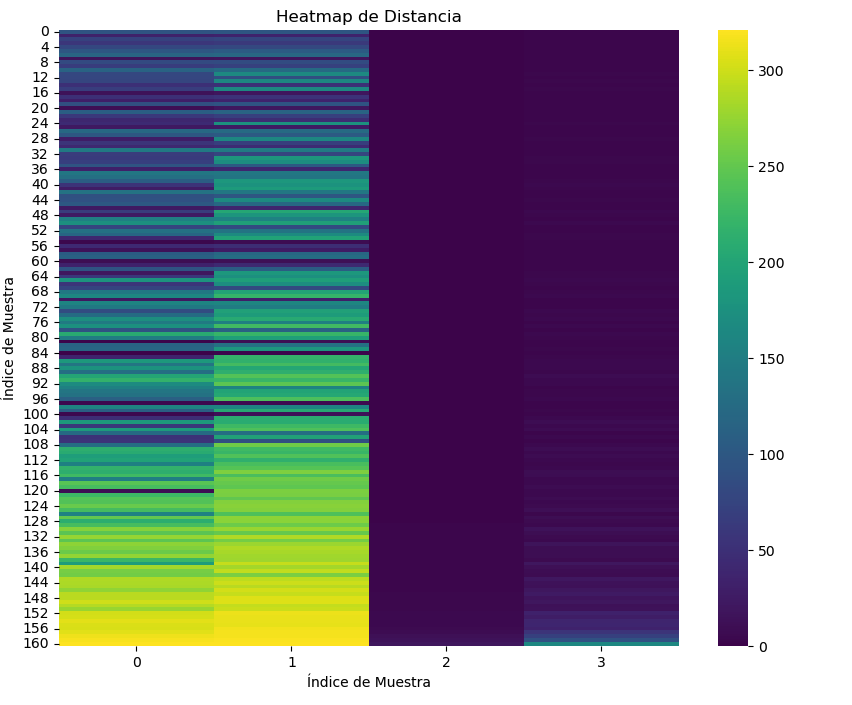
****



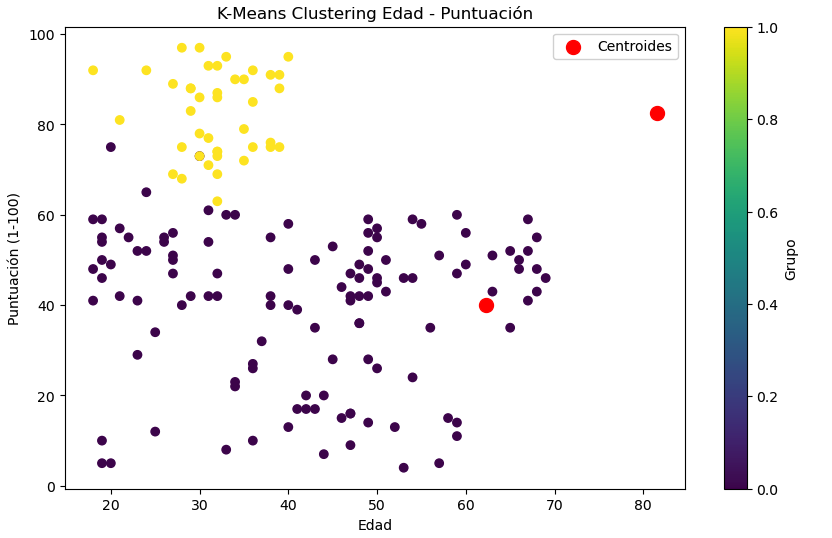
1. **Coeficiente de Silueta (Silhouette Score):** 0.3682282467444023 indica que la separación entre los clusters es moderada. No es excelente, pero tampoco es mala.
2. **Índice de Calinski-Harabasz:** 89.12107986634129 muestra una buena separación y densidad de los clusters. Esto sugiere que los clusters están bien definidos y son distintos entre sí.
3. **Índice Davies-Bouldin:** 0.9818803754390855 sugiere que los clusters tienen alguna superposición o similitud entre sí. Aunque no es muy alto, indica que hay margen para mejorar la separación y la distinción entre los clusters.
4. **Coeficiente de correlación cofenética:** 0.663537885458443 significa que el dendrograma captura de manera razonable las relaciones de similitud entre tus datos originales, pero podría haber margen para mejorar.



Con este grafico se visualiza el detalle del coeficiente de correlación cofenética, como anteriormente nos dio 0.66 acá se confirma y cada punto del grafico hace referencia a una observación de los datos.



El anterior mapa de calor muestra visualmente las distancias entre todas las muestras en tus datos. Las celdas del mapa de calor representan estas distancias, donde colores como el amarillo indican distancias más grandes y colores como la escala de morado indican distancias más pequeñas, como en este caso en el índice de muestra 2 a simple vista se puede identificar distancias 0.



La grafica anterior nos muestra una relación entre la edad y la columna de puntuación, en este caso se seleccionaron 6 grupos y se tienen dos centroides, que son los puntos representativos de cada grupo, marcados de color rojo. En este caso los puntos del grupo 1 tienen una mayor puntuación.