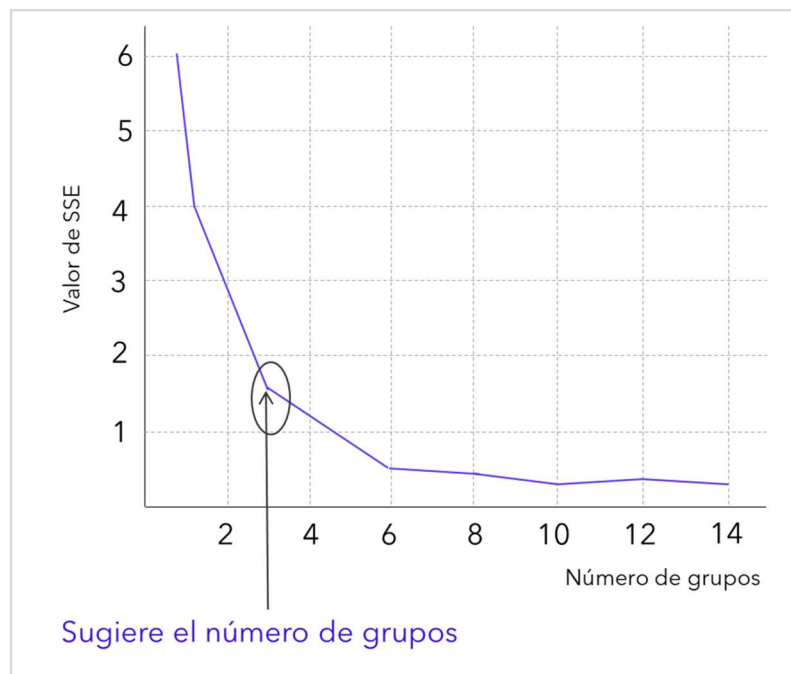


¿Cómo determinar el número de grupos?

En general, los algoritmos basados en centroides, como K-medias y K-medoides, requieren que les especifiquemos el número de grupos (el valor de K) antes de su ejecución. Lo ideal es que el número de grupos lo indiquen los expertos del dominio o que podamos identificar este número con base en las características del problema. Pero, si esto no es posible, ¿cómo determinar el valor de K? Para esto podemos utilizar el método de Elbow y el de la Silueta. Veamos cómo funcionan:

- **Método de Elbow.**

Se basa en la métrica de inercia o suma de las distancias al cuadrado (SSE) para determinar el número de grupos. Para varios valores de K se calcula el valor de esta métrica para la agrupación resultante. Luego se construye una gráfica, como se muestra a continuación:

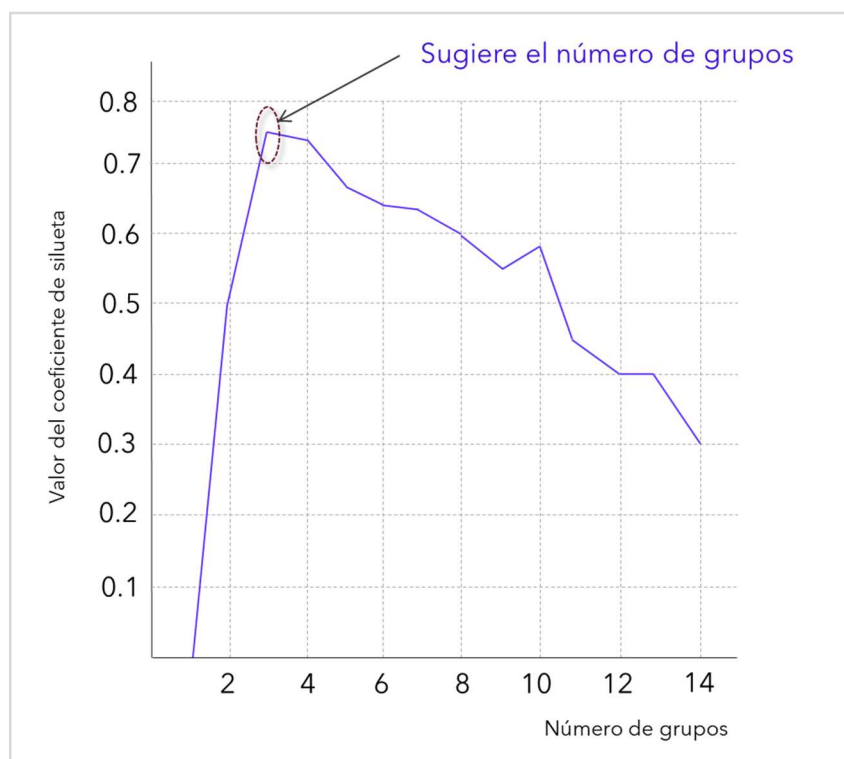


Gráfica 1. Método de Elbow.

En la gráfica se observa un primer cambio brusco en la evolución de SSE, que asemeja un “codo”. Este punto puede dar información del número de grupos para ese conjunto de datos. En este caso, sería $K = 3$.

▪ Método de la Silueta

Este método se basa en la métrica de silueta para determinar el número de grupos. Al igual que en el método de Elbow, para varios valores de K se calcula el valor de esta métrica de rendimiento para la agrupación resultante. Luego, se construye una gráfica, como se muestra a continuación:



Gráfica 2. Método de la Silueta.

En la gráfica se observa cómo evoluciona el valor del coeficiente de silueta a medida que se aumenta el número de grupos. Con $K = 3$ se obtiene el mejor valor de esta métrica, lo que sugiere que este debería ser el número de grupos a utilizar para este conjunto de datos.

Es importante resaltar que estos son métodos heurísticos, lo cual quiere decir que en muchos casos pueden resultar útil y acertar en el número de grupos adecuado para el problema, pero no siempre será así.

Bibliografía

Patel, A. (2019). *Hands-On Unsupervised Learning Using Python. How to Build Applied Machine Learning Solutions from Unlabeled Data*. O'Reilly.

Géron, A. (2022). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. Concepts, Tools, and Techniques to Build Intelligent Systems*. O'Reilly.

© - **Derechos Reservados:** la presente obra, y en general todos sus contenidos, se encuentran protegidos por las normas internacionales y nacionales vigentes sobre propiedad Intelectual, por lo tanto su utilización parcial o total, reproducción, comunicación pública, transformación, distribución, alquiler, préstamo público e importación, total o parcial, en todo o en parte, en formato impreso o digital y en cualquier formato conocido o por conocer, se encuentran prohibidos, y solo serán lícitos en la medida en que se cuente con la autorización previa y expresa por escrito de la Universidad de los Andes.

De igual manera, la utilización de la imagen de las personas, docentes o estudiantes, sin su previa autorización está expresamente prohibida. En caso de incumplirse con lo mencionado, se procederá de conformidad con los reglamentos y políticas de la universidad, sin perjuicio de las demás acciones legales aplicables.
