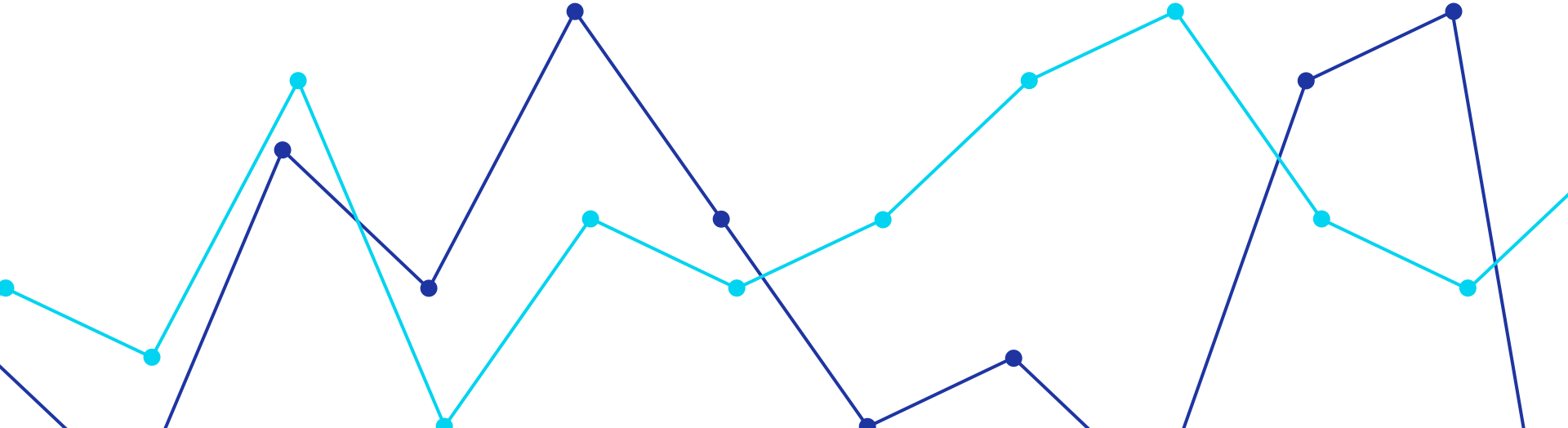


Victor N. Hernández (1301219)

Justificación de implementación

Universidad Rafael Landívar
Análisis de Datos I



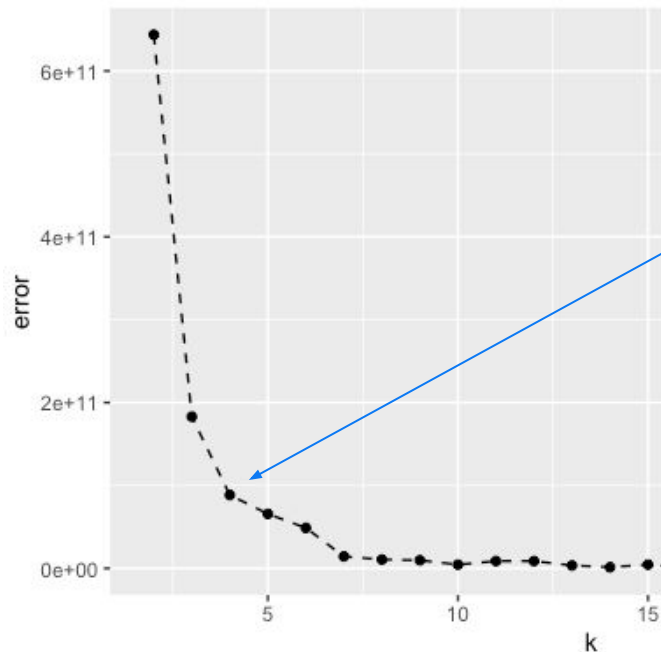
Segmentation data

Proceso elaborado

El análisis realizado, a partir de la data brindada en el dataset “**segmentation data.csv**”, consistió en eliminar preliminarmente las columnas de la información que representan un riesgo de sesgo en el proceso de obtención de la cantidad de clusters indicada para el problema.

Las únicas columnas que se tomaron en cuenta para el análisis son: **Education, Age e Income.**

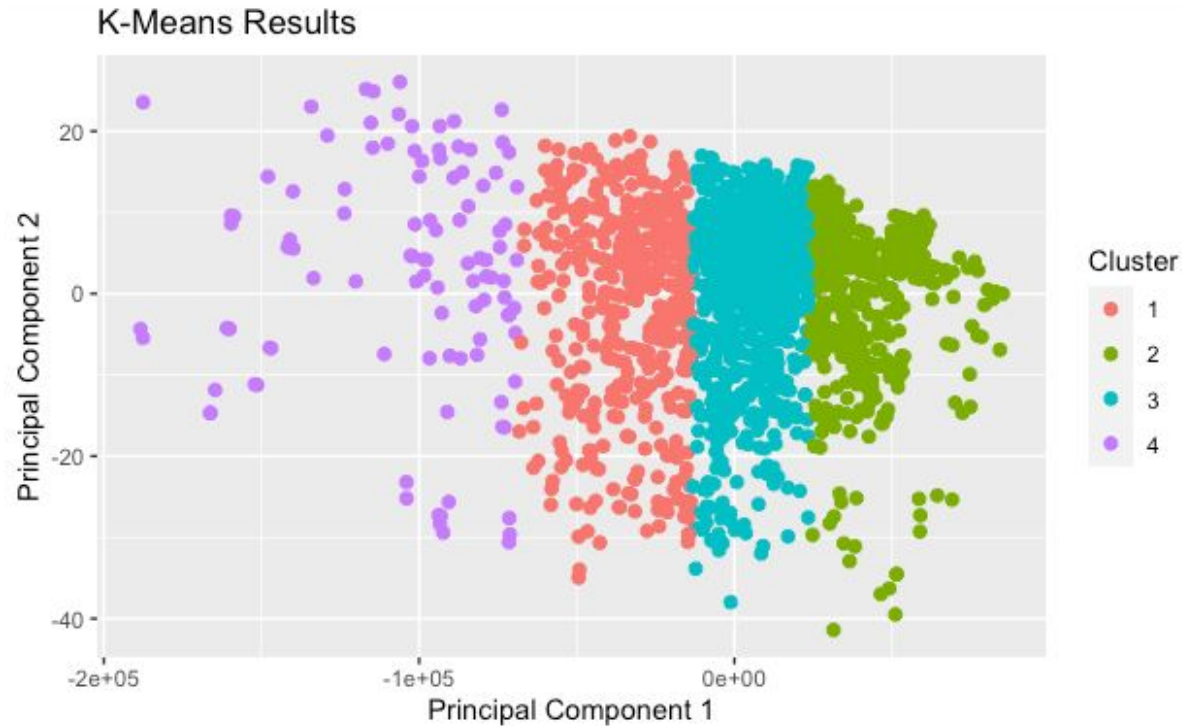
Gráfico resultante



Clusters seleccionados

Con ayuda de la técnica del codo, se determinó que la cantidad de clusters ideal serían 4.

Gráfico obtenido a partir del análisis



Segmentation data

Proceso elaborado

El análisis realizado, a partir de la data brindada en el dataset “**marketing_campaign.csv**”, consistió en eliminar preliminarmente las columnas de la información que representan un riesgo de sesgo (null) en el proceso de obtención de la cantidad de clusters indicada para el problema.

Posteriormente la eliminación, de las columnas que cuenten con valores fragmentados de forma concreta, en este proceso se eliminaron las columnas: “*Kidhome*”, “*Teenhome*”, “*Z_Revenue*”, “*Response*”.

Posterior a esto, se realizó un proceso de análisis basado en el estudio de la frecuencia y densidad de las categorías perteneciente a las columnas faltantes, esto para eliminar el “ruido” en la data, de variables que no representen un impacto real. En este proceso se eliminó de la columna *Marital_Status*, las categorías “*Absurd*”, “*YOLO*” y “*Alone*”.

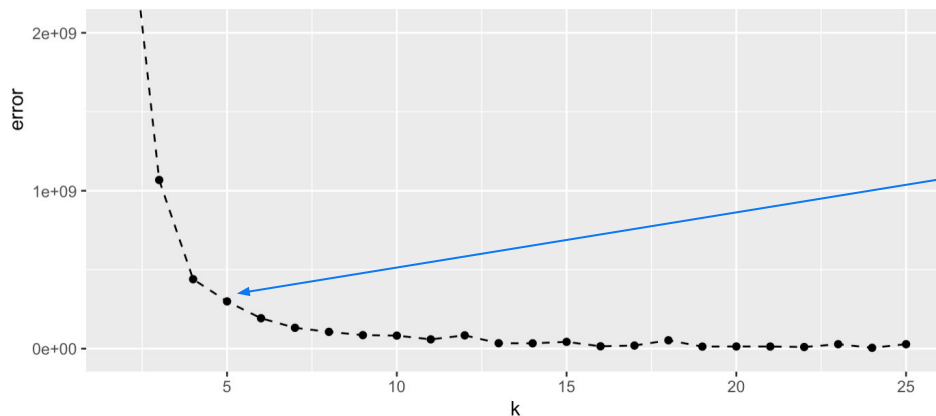
Una vez completado este proceso, se sustituyeron las columnas categóricas “*Education*”, “*Marital_Status*”, “*Dt_Customer*”, a su valor equivalente en materia de densidad en la cantidad total de información. Esto para dejar en nuestro dataset solamente información numérica para el posterior análisis.

Marketing Campaign data

Determinación de clusters

Una vez que se llevó a cabo el proceso de selección y limpieza de la información, se deberá de generar el gráfico para la selección de la cantidad óptima de clusters que se implementarían en el proceso de análisis de la información.

Gráfico resultante



Clusters seleccionados

Con ayuda de la técnica del codo, se determinó que la cantidad de clusters ideal serían 5.

Gráfico obtenido a partir del análisis

