

UNIVERSIDAD RAFAEL LANDÍVAR

FACULTAD DE INGENIERÍA

ANÁLISIS DE DATOS

SECCIÓN 1 VESPERTINA

MGTR. JAIRO OMAR SALAZAR CHÁVEZ

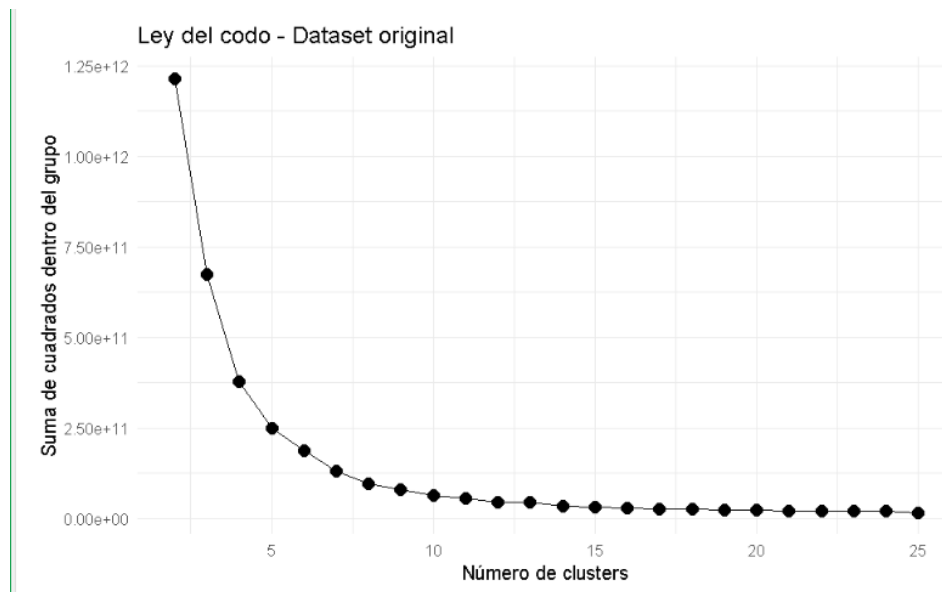
CLUSTERING Y FEATURE ENGINEERING

Julio Anthony Engels Ruiz Coto 1284719

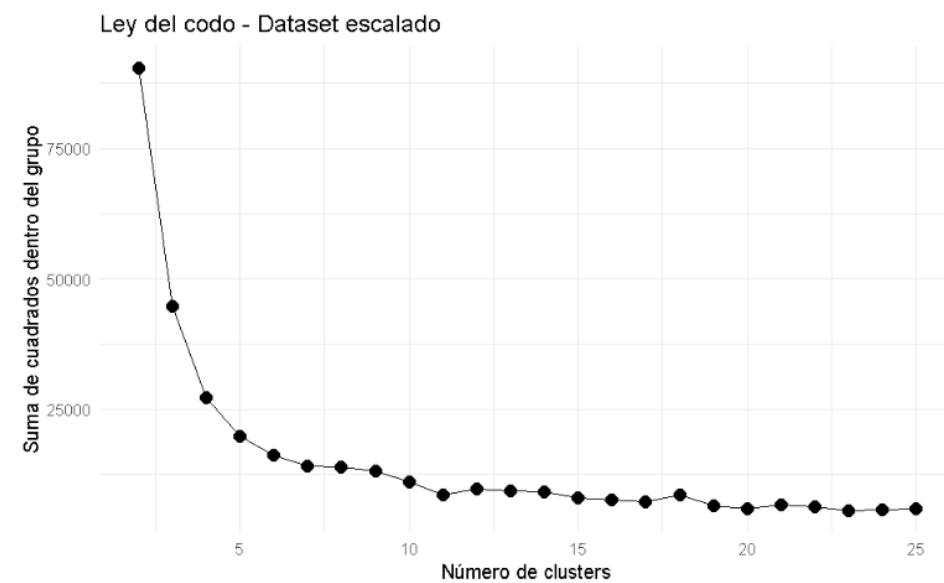
GUATEMALA DE LA ASUNCIÓN, ABRIL 16 DE 2023

CAMPUS CENTRAL

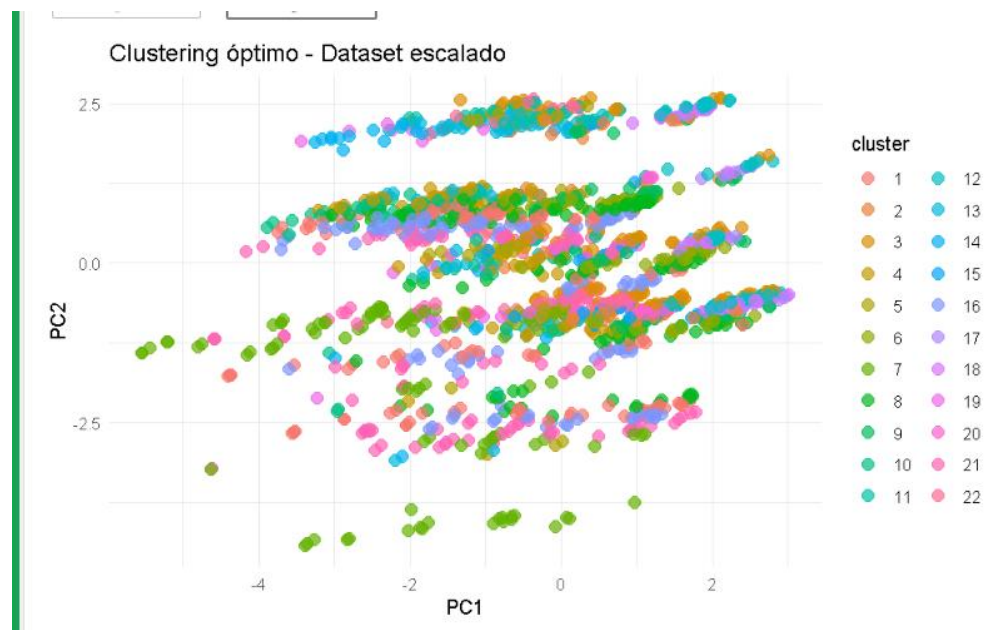
GRAFICA DEL LEY DEL CODO – DATASET ORIGINAL



GRAFICA DEL LEY DEL CODO – DATASET ESCALADO



GRAFICA DE CLUSTERING OPTIMO



PREGUNTAS

1. Existe diferencia en el nivel de errores entre el dataset con el Income original o el dataset con el Income estandarizado? Explique lo encontrado y exponga brevemente la razón principal por la cual existen o no existen diferencias.

Si presenta diferencias en el nivel de errores entre el dataset con Income original y el dataset con el Income estandarizado, lo que puedo observar en la estandarización de la columna Income es que ayuda a que dichas características estén en la misma escala, lo que ayuda a la convergencia de los algoritmos y mejorando la calidad de los clusters, puedo decir que la estandarización permite que todas las características tengan el mismo peso en el proceso de clustering, lo que ayuda a mejorar la precisión de dichos resultados.

2. Existe diferencia en el número de clusters optimos entre el dataset con el Income Original o el dataset con el income estandarizado? Explique lo encontrado y exponga brevemente la razón principal por la cual existen o no existen diferencias.

El número de clusters óptimos depende de la estructura subyacente de los datos y no necesariamente cambiará después de estandarizar la columna Income. Sin embargo, si la estandarización ayuda a revelar patrones en los datos que no eran evidentes antes, podría resultar en un número diferente de clusters óptimos.