

IMPACTO DE LA **CONTAMINACIÓN** EN EL APRENDIZAJE DE LOS ESTUDIANTES CHILENOS

OBJETIVO

Desarrollar un modelo de Clusterización que permita agrupar en base a las características que acompañan a los establecimientos que pertenecen a zonas de sacrificio en Chile.

METODOLOGÍA

1. Análisis exploratorio de datos

- Combinar bases de datos (entre los años de estudio)
- Limpiar base de datos (de valores nulos)
- Redimensionar la base de datos
- Identificar correlación de variables (entre promedios por año)

2. Modelo descriptivo

- Para llevar a cabo el estudio, se utilizó la método KMeans del modelo de clusterización que permite generar grupos representativos de la población de estudio, con diferentes valores para k.

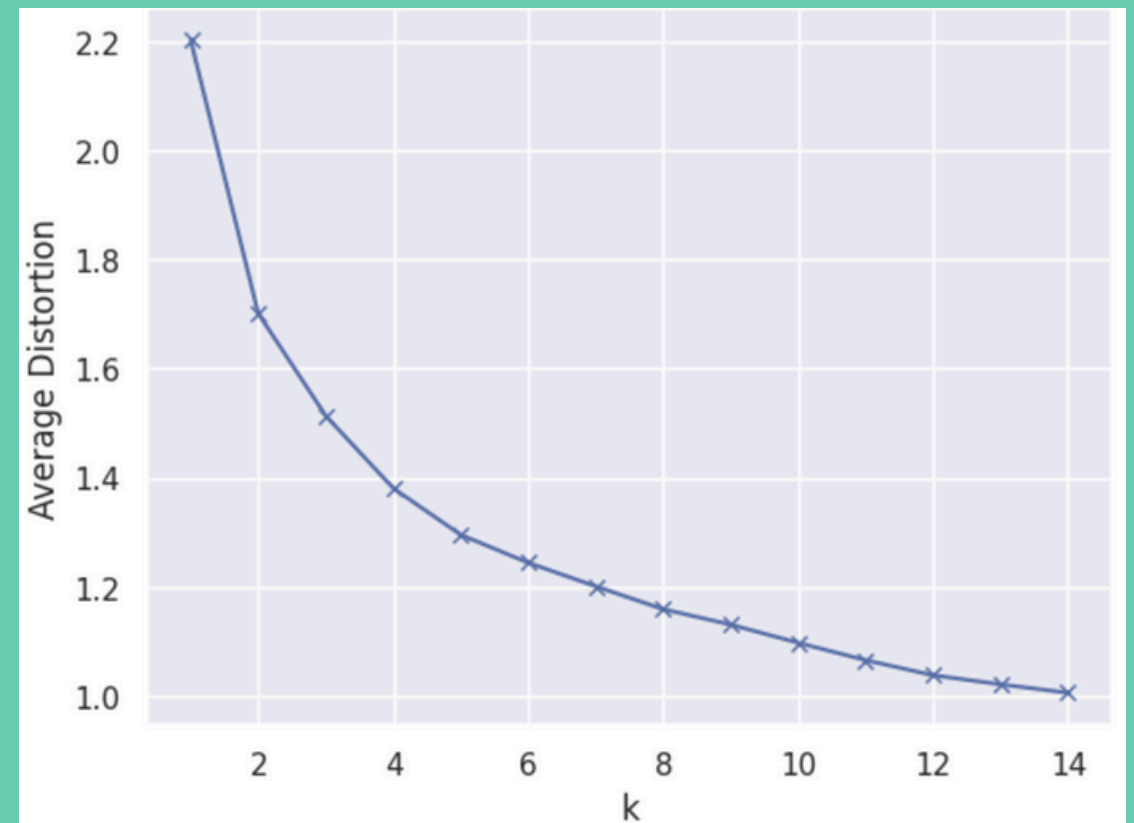
3. Validación del modelo

- Identificación de establecimientos ubicados zonas de sacrificio en cada cluster.
- Visualización de una tendencia de los datos estudiados.

RESULTADOS

Es posible clasificar los cluster entre “bajos”, “medios” y “altos”. Si lo vemos de esta manera, podemos decir que el 83% de los colegios pertenecientes a zonas de sacrificio se encuentran con promedios de rendimiento académico medios y bajos.

Selección de K a través del Método del Codo



Cantidad de colegios en zonas de sacrificio en cada Cluster

KM_segments	Z_Sacrificio	
0	0	783
	1	3
1	0	1689
	1	25
2	0	1526
	1	19
3	0	721
	1	6

Boxplot de los promedios de Lectura y Matemática por Cluster

