

# Task 5

Tema: Clustering  
Profesor: Eduard F. Martínez González

Septiembre 2025

## Introducción

Este taller tiene como objetivo aplicar técnicas de análisis de conglomerados (clustering) para segmentar las entidades financieras reportadas por la Superintendencia a septiembre de 2024. La base de datos suministrada contiene información sobre `codentidad`, `nombre de la entidad`, `activos` e `ingresos`.

El ejercicio busca identificar patrones en el comportamiento de las entidades y clasificarlas en tres grupos homogéneos, lo cual permitirá interpretar diferencias relevantes entre entidades con distintos niveles de tamaño y desempeño financiero.

## Instrucciones

- El acceso al repositorio del taller está disponible en el siguiente enlace:
- La entrega debe incluir:
  - Un archivo en formato `.R` con el desarrollo del taller.
  - Un archivo en formato PDF con las respuestas e interpretaciones solicitadas.
- El script en R debe contener:
  - Nombres de los integrantes y versión de R utilizada.
  - Carga de librerías necesarias (`dplyr`, `ggplot2`, `stats`, etc.).
  - Código organizado y documentado con comentarios: `## Punto 1`, `## Punto 2`, etc.

```
## Nombre(s) de Autor(es)
## R version 4.5.0

rm(list=ls()) # limpiar entorno

## cargar librerias
```

```
require(dplyr)
require(ggplot2)
require(janitor)
require(cluster)

# cargar datos
df <- read.csv("data_clustering.csv")

# Vista previa
glimpse(df)
```

## Taller: Clustering de Entidades Financieras

1. (15 pts) Revise la base de datos: verifique la variable `codentidad`, identifique CDACs y realice limpieza de datos. Elimine observaciones atípicas en activos e ingresos usando el criterio de Tukey.
2. (25 pts) Estandarice las variables `activos` e `ingresos` y realice un `k-means` con  $k = 3$ . Reporte el tamaño de cada cluster.
3. (20 pts) Genere un gráfico de dispersión con `activos` en el eje X e `ingresos` en el eje Y, coloreando los puntos por cluster.
4. (20 pts) Presente estadísticas descriptivas (promedio, mediana, mínimo, máximo) de activos e ingresos para cada cluster.
5. (20 pts) Interprete los resultados: describa las características principales de cada grupo (entidades grandes, medianas y pequeñas) y comente qué hallazgos resultan más relevantes para el análisis financiero.

**Total: 100 pts**