**SESIÓN 6: ANALISIS DE COMPONENTES PRINCIPALES**

Disponemos de una muestra de 2000 clientes de una entidad bancaria, dividido en 1000 que fueron baja y 1000 que no lo fueron. Los datos se encuentran en el fichero “churn.txt”.

El objetivo es comprobar la homogeneidad de los dos conjuntos de clientes (bajas o no) respecto de su posición bancaria habitual (un año antes de la potencial baja). Y en segundo lugar queremos obtener una tipología de clientes según su posición bancaria (esto se obtendrá en el siguiente ejercicio).

La posición bancaria de un cliente viene definida por el saldo en cada uno de los productos bancarios de pasivo y activo. En nuestro caso esta información viene reflejada de forma agregada en las siguientes variables:

[13] "Total\_activo" "Total\_Plazo" "Total\_Inversion"

[16] "Total\_Seguros" "Total\_Vista"

Para ello realizaremos un Análisis de Componentes Principales, para ver cuáles son los factores latentes que estructuran los datos y minimizar la parte de fluctuación aleatoria para a continuación efectuar el “Clustering” (en la 7ª sesión).

1. Lea el fichero “churn.txt”. Efectúe las correcciones del fichero de datos ya realizadas para el ejercicio de “Profiling”. Razone lo que significa realizar un ACP con datos estandarizados o sin estandarizar.
2. Efectúe un Análisis de Componentes Principales estandarizado, tomando como variables activas los productos bancarios antes especificados. Declare como ilustrativos los clientes que se dieron de baja (de esta forma la configuración obtenida reflejará la de los clientes “normales”).
3. Obtenga la representación gráfica del ”Screeplot” (diagrama de los “eigenvalues”) y a la vista de las correlaciones entre las variables originales y las componentes principales, decida el número de dimensiones significativas. ¿Cuál es el porcentaje de variancia retenido?.
4. Efectúe una rotación “varimax” para hacer más evidente los factores latentes (intangibles) presentes en sus datos activos. ¿Cuáles son en este caso estos factores latentes?.
5. Represente gráficamente la nube de puntos individuo activos. Sobre esta nube proyecte los individuos suplementarios. ¿Piensa Ud. que la *configuración* de los clientes que han sido baja es distinta de la de los clientes que no han sido baja?. Haga la representación solamente en el primer plano factorial (ejes 1 y 2).