# Silicon Valley Diversity Data

## La diversidad en 23 empresas top en tecnología

## Objetivo

La finalidad de este documento es la de demostrar los conocimientos aprendidos en la asignatura y comprobar los niveles de diversidad y diferencia laboral entre hombres y mujeres en 23 empresas diferentes de Silicon Valley.

## Dataset

El conjunto de datos escogido representa a los empleados de las 23 diferentes empresas.

Entre los atributos encontramos la empresa en la que trabajan, su raza, su sexo y su puesto de trabajo dentro de la empresa. También aparece el año de los datos, pero nos es irrelevante a la hora de trabajar con los datos, y el número de personas que trabajan en la empresa con esa combinación de raza, sexo y puesto.

Aparte de los datos de los trabajadores, también tenemos datos totales sobre todas las combinaciones posibles de empresa, raza, sexo y puesto de trabajo.

## Dataset procesado

El dataset ha tenido que ser modificado por mi ya que era necesario para poder operar con los datos de la manera que yo quería.

En primer lugar, he eliminado el atributo del año, ya que, como he dicho antes, me era irrelevante porque siempre era el mismo valor.

Después he filtrado las filas de los datos totales porque tampoco los necesitaba.

Y en ultimo lugar, he escrito un programa en Java que en función del valor del atributo count, que contaba el número de personas con cada combinación de raza, sexo y puesto de trabajo, escribe las combinaciones tantas veces como aparecía en la columna count.