



ACTIVIDAD

Ejercicio de variables

Los archivos CSV, o valores separados por comas, son un tipo de archivo de texto que se utiliza para almacenar datos en forma de tabla. Cada fila del archivo representa una fila de la tabla, y cada columna se separa por una coma. Estos archivos se pueden visualizar con múltiples programas, por ejemplo, con un editor de texto (block de notas) en el cual se ven como en la figura 1:

```
Nombre,Apellido,Grado  
Juan,Perez,10  
Maria,Gonzalez,11  
Pedro,Sanchez,12
```

Figura 1. Visión de un archivo CSV.

En el archivo de la figura 1, cada fila representa un estudiante. La primera columna contiene el nombre del estudiante, la segunda columna contiene el apellido del estudiante, y la tercera columna contiene el grado del estudiante.

Los archivos CSV son populares para almacenar conjuntos de datos ya que son fáciles de crear, fáciles de leer para máquinas y para humanos y se pueden editar en cualquier procesador de texto. También son compatibles con muchos programas de hojas de cálculo y son eficientes en almacenamiento de espacio.



Para este ejercicio se requiere acceder al dataset del titanic (<https://github.com/datasciencedojo/datasets/blob/master/titanic.csv>). Este dataset es un archivo .csv que contiene la información de las personas que iban a bordo del titanic e información acerca de su supervivencia.

Se plantea realizar las siguientes actividades con los estudiantes:

1. Abrir el archivo con un sistema de hojas de cálculo (Excel o Google sheets).
2. Visualizar las columnas que tiene el archivo y, con ayuda de internet determinar qué significa cada una de ellas.
3. Cambiar el color de cada una de las columnas dependiendo del tipo de variable que tienen, (cualitativa o cuantitativa) según se aprendió en la lección.
4. Calcular el promedio de supervivencia general
5. Calcular el promedio de supervivencia de los hombres
6. Calcular el promedio de supervivencia de las mujeres
7. Según la cantidad de clases de tiquete, calcular el promedio de supervivencia para cada uno de ellos.

Una vez se realicen los siete puntos descritos, se debe discutir con los estudiantes los resultados observados y cómo cada categoría influye en la supervivencia.