

## ACTIVIDAD 1

### Tipo actividad: Ejercicio de revisión de datos (2 horas)

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 (link en el cuadro de la actividad). Este dataset es un archivo .csv que contiene la información de las personas que iban a bordo del titanic.

Como ejercicio se debe abrir el archivo .csv utilizando un programa de texto como el block de notas de Windows, un editor de texto como Vim o como visual studio code.

Se plantea realizar las siguientes actividades con los estudiantes:

1. En el archivo de texto, determinar cuántos registros tiene el archivo
2. ¿Es posible saber cuántos pasajeros subieron desde esta forma de ver el archivo?
3. Pida a los estudiantes que lean la primera fila del archivo csv, e intenten establecer qué significa cada una de ellas.
4. Posteriormente, oriente a los estudiantes a buscar el significado de cada una de las columnas.
5. Finalmente Discuta con los estudiantes si les parece fácil la visualización del archivo a través de un editor de texto.

Como segunda parte de la actividad se establecerá un contraste entre usar una herramienta como un editor de textos para ver un conjunto de datos y el uso de software de hojas de cálculo. Para esto, deberán abrir una nueva hoja de cálculo en Google sheets (también es posible usar Excel o

libre office calc). Para ir a google sheets diríjase a la dirección <https://docs.google.com/spreadsheets> y cree una hoja de cálculo nueva. Allí en el menú archivo haga click en importar y cargue el fichero .csv con los datos del titanic, se debe ver como la imagen de la figura 2.



Figura 2: Menú de importar archivos.

Al hacer click en el botón importar datos, se deben ver la tabla como la de la figura 3.

PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	Cabin	Embarked
1	0	3	Brand, Mr. Owen	male	22	1	0	A5.21171	7.25		S
2	1	1	Cummings, Mrs. J.	female	38	1	0	PC 17599	71.2833	C85	C
3	1	3	Hekkersen, Mrs.	female	26	0	0	STON/O2. 31012	7.925		S
4	1	1	Futrelle, Mrs. J.	female	35	1	0	113803	53.1	C123	S
5	0	3	Allen, Mr. William	male	35	0	0	373450	8.05		S
6	0	3	Moran, Mr. James	male		0	0	330877	8.4503		Q
7	0	1	McCarthy, Mr. Thomas	male	54	0	0	17463	51.8625	E46	S
8	0	3	Palsson, Master.	male	2	3	1	349909	21.075		S
9	1	3	Johnson, Mrs. O.	female	27	0	2	347742	11.1333		S
10	1	2	Nasser, Mrs. Nicola	female	14	1	0	231736	30.0700		C
11	1	3	Sandstrom, Miss.	female	4	1	1	PP 9549	15.7	06	S
12	1	1	Bonnell, Mrs. E.	female	58	0	0	113783	26.56	C103	S
13	0	3	Saunderscock, Mr.	male	26	0	0	A5. 2151	8.05		S
14	0	3	Anderson, Mr. F.	male	39	1	5	347082	31.275		S
15	0	3	Vestrom, Mrs. F.	female	14	0	0	350406	7.8542		S
16	1	2	Hewlett, Mrs. J.M.	female	55	0	0	248706	16		S
17	0	3	Rice, Master. Eu.	male	2	4	1	30262	29.125		Q
18	1	2	Williams, Mr. Ch.	male		0	0	244373	13		S
19	0	3	Yardley Blake, F.	female	31	1	0	345753	18		S
20	1	3	Masseboom, Mr.	male		0	0	2649	7.225		C
21	0	2	Fynney, Mr. Jose	male	35	0	0	236665	26		S
22	1	2	Beesley, Mr. Law	male	34	0	0	240696	13.066		S
23	1	3	McCowan, Miss.	female	15	0	0	330923	8.0202		Q
24	1	1	Sloper, Mr. Willie	male	28	0	0	113788	35.5	A6	S
25	0	3	Palsson, Miss. T.	female	8	3	1	349909	21.075		S
26	1	3	Asplund, Mrs. G.	female	38	1	5	347077	31.3875		S
27	0	3	Eric, Mr. Fareed	male		0	0	2631	7.225		C
28	0	1	Fortune, Mr. Chas	male	19	5	2	19650	263	C25 C26 C27	S
29	1	3	O'Day, Miss. F.	female		0	0	330669	7.8702		Q
30	0	3	Todoroff, Mr. Luke	male		0	0	349216	7.8956		S

Figura 3: Hoja de cálculo con la información del titanic.

Repita el ejercicio de pedir a los estudiantes enumerar la cantidad de registros (pasajeros en el viaje) y posteriormente pida a los estudiantes que encuentren quien es el pasajero con mayor edad a bordo.

Pregunte a los estudiantes qué tan fácil sería buscar ese pasajero empleando la vista del CSV con un editor de texto.

Finalmente pida a los estudiantes que determinen cuantos sobrevivientes hubo de total de pasajeros y saque conclusiones del uso de una hoja de cálculo en contraste con un editor de texto.