

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 clic en importar y cargue el fichero .csv con los datos del titanic, se debe ver como la imagen de la figura 2.

Importar archivo		×
Archivo		
titanic.csv		
Importar ubicación	Tipo de separador	
Reemplazar hoja de cálculo 👻	Detectar automátic	amente 👻
Convertir el texto en números	, fechas y fórmulas	
	Importar datos	Cancelar

Figura 2: Menú de importar archivos.

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





						- Sec														
						ΤΙ														
-	Hoja de cálo	culo sin titul	lo 🕆 Ba d	2																
	Archivo Edite	ar Ver Inse	ettar Formatio	Datos Herra	mientas Extern	siones Aya	ada													
	A Meniis 🔊 😏	2 O S	100% •	s × n	0g t23 Pred	e. • -	1	0 +	8	I - A	۰,	⊞ 8 .	1E *	<u>+</u> =	м - 7	5 m	60 J	1 10	Y B	ē
	- 1 A P	and the second second																		
10	• 1 Ja •	corriger to						-								-				
	A		G	0	E			0		1	-					κ.	-	-		
-	Passergend 13	surved.	POIdes	Name	50K	Age	-	5460		Pach	10	6.5	Fare		Cabin		the	arked		
÷	1	0		fibund, bir Ow	1 mase		22		-1		0 40	621121		7.25	100		- 2			
-	2			Carnings, Mrs.	a lemale		~		1		0 PG	1/009		11.2535	080					
÷	3		-	Freikolenen, Miss	Retruike		2				0 50	ALCORE STORES	-	1.560			2			
-	4	1		Public, Mrs. Ja	s enale		-20		1			123860		50.1	UT25		8			
÷	5	0		bloom Mr. Inter	r mare		20		0		0	330077		0.4500	-		-			
-	7	0		Marcada, Mr. Jana	e male		54		0		0	17863		61 9806	5.40		0			
		0		Dataset Maria	r main		2					140000		21.075			1			
10	9			inhoson Mrs. (Imple		22		0		2	347742		11 1333	_		8			
11	10		2	Nexuer Mex. No.	- Arrule		14		1		a	230736		30.0708			c			
12	11	1	3	Sandstrom, Mis	e female		4		1		1 PP	9549		95.7	06		5			
10	12	1	1	Borwit Mos. 8	Arrule		58		0		0	113763		26.55	C108		8			
14	13	0		Saudercock, M	6 maile		20		0		0 4/5	2151		8.05			5			
15	14	0	3	Andersson, Mr.	/ male		39		1		5	347062		31.275			8			
15	15	0	1 3	Vedrare, Men.	l fernalie		14		D		a	350406		7.8542			s			
17	16	1	2	Hewlett, Mrs. (h	t temale		55		0		0	243706		16			8			
18	17	0	3	Res, Mader D	a mailer		2		- 4		1	302852		29.125			Q			
19	18	1	2	Williams, Mr. O.	hmale				0		0	244373		13	i		8			
20	19	0	3	Vanie Planke,	5 formalie		31		1		0	345/63		16			8			
21	20	1	9	Masselmani, M	e fernale				0		0	2649		7.225	i		G			
22	21	0	2	Fyrney, Mr. Jas	(male		35		0		0	230865		26			8			
23	22	1	2	Beesley, Mr La	u maie-		34		D		a	243656		13	055		5			
24	23	1	3	McGowan, Miss	, lenale		15		0		0	330923		8.0292			Q			
25	24	1	1	Sloper, Mr. Will	e made		28		D		0	112/16		35.5	AB		5			
29	25	0		Palsson, Miss.	ternate				3		1	349909		21.076	_		8			
217	26		3	Appland, Mrs. C	A NOTING		38		1		9	34/0/7		a1.36/b	_		5			
20	20	0		Erw, W. Fared	mae				0			2531		128			G			
20	28	0		Citrane, Mr. Ch	1 TAC		12		3		6	10050		263	030	sun	3			
14	29	1		Tobal Inc.	e terrate				0		0	3,400-40		2 8000	_					
	30			AND A DECK OF A LOD	The second se				0		u	24.50 13		1.00000			3			

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 cuántos sobrevivientes hubo de total de pasajeros y saque conclusiones del uso de una hoja de cálculo en contraste con un editor de texto.



