**TALENTO TECH**

**ANALISIS DE DATOS NIVEL INTEGRADOR G2**

**ACTIVIDAD 1 – LECCIÓN 1**

**ESTUDIANTE:** Andrés Fernando Ruales Muñoz

**PROFESOR:** Andrés Felipe Escallón Portilla

**POPAYÁN – CAUCA**

**2024**

**ACTIVIDAD PROPUESTA**

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. Finalmente Discuta con los estudiantes si les parece fácil la visualización del archivo a través de un editor de texto.
5. 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 quién es el pasajero con mayor edad a bordo.
6. 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.

**SOLUCIÓN**

1. El dataset del titanic tiene en total 891 registros.
2. Si es posible conocer este dato gracias a que en la columna “PassengerId” es incremental en 1 por cada pasajero.

* **PassengerId**: Identificación del pasajero.
* **Survived**: Indica si el pasajero sobrevivió (1) o no (0).
* **Pclass**: Clase del boleto (1 = Primera clase, 2 = Segunda clase, 3 = Tercera clase).
* **Name**: Nombre del pasajero.
* **Sex**: Género del pasajero.
* **Age**: Edad del pasajero.
* **SibSp**: Número de hermanos/cónyuges a bordo.
* **Parch**: Número de padres/hijos a bordo.
* **Ticket**: Número de ticket.
* **Fare**: Tarifa del pasaje.
* **Cabin**: Número de cabina.
* **Embarked**: Puerto de embarque (C = Cherbourg, Q = Queenstown, S = Southampton).

1. Desde un editor de texto como el bloc de notas es difícil ya que al no tener un orden visual se dificulta poder realizar un análisis a las columnas y poder encontrar relaciones que tengan sentido según lo que queramos encontrar.

* El total de pasajeros es de 891.
* El pasajero de mayor edad es Connors, Mr. Patrick, que tiene una edad de 70.5 años.

1. Seria un poco complicado teniendo en cuenta que el dataset del titanic es demasiado grande y tocaría buscar manualmente el pasajero que queremos.
2. El numero total de sobrevivientes del Titanic fue de 342 personas.

* Utilizando un editor de texto para realizar esta tarea sería mucho más difícil. Ya que tocaría revisar manualmente cada fila del dataset y buscar la información relevante sobre la supervivencia de cada pasajero, lo que sería muy tedioso en cambio al utilizar una hoja de cálculo, se puede obtener el número total de sobrevivientes de manera rápida y precisa ya que manipular y analizar datos en función de la tarea específica que se esté realizando.