

ACTIVIDAD #1

Tipo actividad: Hoja de cálculo de exploración

Actividad 1. Duración: 2 horas

Antes de realizar tareas de minería y analítica de datos, es apropiado conocer el contenido de una base de datos, sus variables, las relaciones entre las columnas y el significado o proveniencia de cada una. De esta forma es más fácil comprender cómo puede llevarse a cabo el preprocesamiento de los datos y conocer qué sucesos están explicando los datos. La idea del proceso de exploración inicial es determinar el contenido de una base de datos y validar o rechazar los aspectos preconcebidos acerca de estos.

Para la actividad los estudiantes deben abrir la hoja de datos de Google sheets que contiene la información de los registros del barco Titanic. El objetivo principal es guiar a los estudiantes a realizar un análisis exploratorio para luego explicar los conceptos del mismo.

Pasos propuestos a los estudiantes para que realicen con una hoja de cálculo la exploración:

1. Identificar cuantas columnas hay y qué significa cada una de las columnas, guíe a los estudiantes a explicar qué significa que exista un número 0 o 1 en la columna “siblings”, o en la columna titulada Parch.
2. Pregunte a los estudiantes cuáles columnas permiten realizar cálculos matemáticos sobre ellas.
3. Indague con los estudiantes, qué significa la letra en la columna embarked.
4. Pregunte a los estudiantes qué significa cada fila en la tabla ¿hacer referencia a una persona, a una cabina, a un puerto o a un tiquete?
5. ¿Qué significa el contenido de la columna Ticket?

Hasta este punto se tiene un análisis previo de los datos, en el cual se conocen los aspectos básicos del conjunto de datos. Cada estudiante debe saber claramente qué significa cada columna, qué información guarda y cómo se debe interpretar.

Para finalizar la actividad oriente a los estudiantes a reemplazar las letras en la columna embarked por su significado para que sea más legible la información.

Pregunte a sus estudiantes ¿Qué otros reemplazos podrían mejorar la lectura de los registros?