

## Contexto

Los temas más importantes vistos en las clases de R, fueron la manipulación y visualización de datos con Tidyverse. Porque en un principio lo que se espera de un analista o científico de datos, es que sea capaz de tomar cualquier conjunto de datos para manipularlo y visualizarlo a su gusto. Al final de este proceso es cuando se extrae la información más importante y luego se comparten los resultados con el resto de la organización.

## Objetivo

Realiza un análisis exploratorio en R, del conjunto de datos Titanic para comprender mejor qué características (por ejemplo, edad, nivel socioeconómico, costo de ticket, puerto de embarcación, etcétera) tienen los pasajeros que sobrevivieron o no sobrevivieron al evento. Una vez finalizado el análisis, en un documento pdf comparte un resumen no mayor a 2 párrafos con tus descubrimientos más importantes.

## Detalle de las columnas

- PassengerId - Identificador único del pasajero.
- Survived - Indicada si el pasajero sobrevivió o no (Yes, No).
- Pclass - Nivel socioeconómico (Upper Class, Middle Class, Lower Class).
- Name - Nombre del pasajero.
- Sex - Género del pasajero.
- Age - Edad del pasajero.
- SibSp - Número total de herman@s y espos@s del pasajero.
- Parch - Número total de padres e hijos del pasajero.
- Ticket - Número de ticket del pasajero.
- Fare - Costo del ticket.
- Cabin - Número de cabina del pasajero (puede contener caracteres vacíos).
- Embarked - Puerto de embarcación del pasajero.

## Pasos a seguir

1. Importa la base de datos *Titanicv2.csv* en R.
2. Realiza un análisis exploratorio de los datos con Tidyverse.
  - Esto incluye el uso de funciones select, filter, summarise, groupby con el operador pipe >%> para conseguir cuentas totales, promedios, medianas, máximos o cualquier otra métrica que consideres pertinente.
  - Esto incluye el uso de visualizaciones como histogramas, diagramas de caja, diagramas de dispersión o más, para identificar patrones, grupos, o tendencias.
  - Utiliza las columnas que tu consideres importantes. ¡No tienen que ser todas! Confía en tu instinto - sean 2, 3 o todas las columnas las que analizaste... no afecta en tu calificación.
3. Guarda tu script de R, pues es parte de la evidencia del caso práctico.

4. Realiza un resumen no mayor a 2 párrafos con los descubrimientos más importantes en un archivo PDF.
5. Guardar el script de *R* con el código que utilizaste para explorar los datos, y el resumen en PDF. Por último sube ambos archivos en la plataforma de la Universidad de Colima.

## **RECUERDA**

Nadie conoce mejor los datos que quien los manipula. Cuando uno es analista o científico de datos debe confiar en sus habilidades analíticas, y sentirse seguro de que el camino que tomó para su análisis es el mejor. En este caso práctico no hay respuesta correcta, ÚNICAMENTE importa cómo haces uso de *R* y Tidyverse para realizar tu análisis exploratorio.