

## Práctica Módulo 3

# R Programming

En base a las tablas facilitadas en formato “csv” dentro de la ruta “m3-rprogramming\_eda\data\wine\_quality” importar las tablas en Rstudio y realizar los siguientes ejercicios:

### Ejercicio 0:

Limpiar los nombres de ambas tablas y pasarlos a ‘snake\_case’, identificar las formas (cantidad de filas y columnas) de cada tabla y verificar que los nombres de las variables coinciden. Por último mostrar 5 observaciones al azar de cada dataframe.

### Ejercicio 1:

- A. Adicione una columna con el nombre ‘class’ a cada uno de los dataframes. En el de vinos tintos deberá tener para todas las observaciones el valor ‘red’ mientras que para vinos blancos deberá tener el valor ‘white’.
- B. Unifique ambas tablas en un sólo dataframe pegando las filas de uno por debajo del otro. Puede resultar de utilidad explorar la función “rbind” de las funciones básicas de R.

### Ejercicio 2:

- A. Analice la estructura y en base al resumen del dataframe resultante determine si los tipos de variables se encuentran asignados correctamente.
- B. En caso de encontrar variables codificadas como character que debieran ser numéricas o variables de tipo character que tendría sentido que sean factores, realizar las operaciones de coerción de tipos según corresponda.

- C. Revisar nuevamente la estructura de los datos corregidos y realizar un resumen de los datos.

### Ejercicio 3:

- A. Compute la matriz de correlaciones empleando la correlación de Pearson para cada una de las combinaciones bivariadas en los datos numéricos empleando observaciones completas y determine las mayores correlaciones positivas y negativas. Qué sugieren?
- B. Repita el punto anterior separando el dataframe limpio en vinos blancos y tintos, generando 2 matrices de correlaciones. Encuentra diferencias que le llamen la atención?

### Ejercicio 4:

- A. Desarrolle una función que permita identificar las observaciones con valores extremos de la variable alcohol a partir del criterio del IQR. (difícil)
- B. En base a los valores que toma la variable hay algo que le llame la atención? Los datos le parecen irregulares? Puede determinar el origen del problema?
- C. Importar nuevamente los datos empleando la función `read_delim()` y ver un summary del dataset completo