

Universidad Nacional Autónoma de México
Facultad de Ciencias



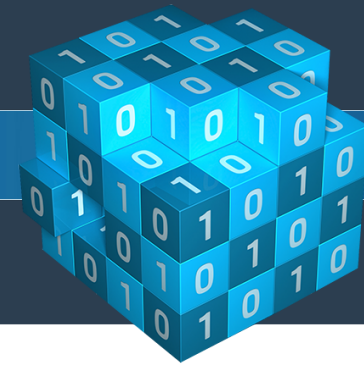
Profesores:

Dra. Amparo López Gaona
M. I. Gerardo Avilés Rosas

Laboratorista:

José Luis Vázquez Lázar





1. Objetivo

Al finalizar esta práctica el alumno será capaz de:

- Apreciar las distintas problemáticas de un conjunto de datos y sus implicaciones.
- Conocer la importancia de la limpieza de datos y aplicar soluciones a problemas de calidad de datos.

2. Marco teórico

El **Preprocesamiento de Datos** engloba a todas aquellas técnicas de análisis de datos que permite mejorar la calidad de un conjunto de datos de modo que las técnicas de extracción de conocimiento/minería de datos puedan obtener mayor y mejor información.

Existen varias técnicas que nos permiten hacer **Preprocesamiento de Datos**:

- Limpieza de datos.
- Integración de datos.
- Reducción de datos.
- Transformación de datos.
- Discretización de datos.

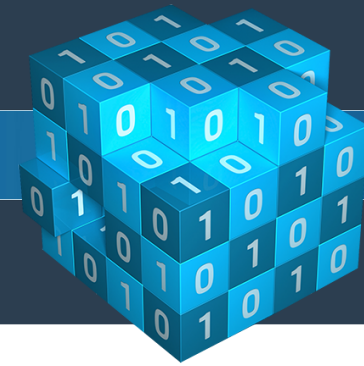
3. Instrucciones

- Descargar e instalar el software **RStudio**
- Retoma el archivo **Adult.csv** de la práctica pasada.

4. Actividades

1. Para cada columna del archivo **Adult.csv**, asigne un nombre en español.





2. Ya identificó los valores ausentes en cada atributo. Asigne un valor a los mismos tanto en el caso numérico como en el caso nominal. Explique que desventajas tiene el método que elige para la imputación, ¿cuál sera uno mejor? y ¿por qué? Haga esto con la ayuda de **RStudio**.
3. En el atributo correspondiente al **estado civil**, la administración no está conforme con que los valores estén en inglés por lo cual se te solicita que los remplace por valores en español. Haga esto en un proceso. Haga esto con la ayuda de **Rstudio**.
4. No olvide que todos sus procesos que modifiquen el conjunto de datos deberán arrojar un nuevo archivo con las modificaciones.

5. Entregables

Deberás enviar un archivo .zip, con nombre <número de cuenta>_practical0, que contenga lo siguiente:

- El proyecto RStudio con las actividades solicitadas.
- Un archivo README.txt que contenga tu nombre completo, tu número de cuenta y tu correo.

a la dirección de correo luiz_lazaro@ciencias.unam.mx con el asunto [A&MD2018-2]Practical0.

