

TPN°2: Análisis Exploratorio y Curación de Datos

Mentoría 10: ¿Cómo vivimos las y los argentinos? Relación entre hábitos de vida, alimentación y prevalencia de las principales enfermedades no transmisibles.

Una vez que las tareas de visualización de datos han concluido (aunque en general siempre continúan) y que uno se encuentra familiarizado con su dataset es importante realizar la curación del mismo. En general, en los diferentes set de datos existen variables con muchos valores espurios o nulos o valores atípicos en distintas variables que es importante atender (entre otros problemas que pudieran aparecer). No solo para la correcta interpretación de los mismos, sino porque el objetivo final puede ser construir un modelo predictivo que se vería afectado por estos problemas en las variables. Por ello, en el presente práctico puede resultar interesante la curación de las variables que puedan resultarles de interés para incorporar en un futuro modelo capaz de predecir la posibilidad de poseer algún tipo de enfermedad no transmisible (en función también de lo que estuvieron realizando en el práctico anterior).

Visualice nuevamente la base de datos utilizada en el trabajo práctico anterior, con la experiencia adquirida y realice un análisis y curación, que podría implicar las siguientes tareas:

- Fusión de conjuntos y/o registros de datos, creación de nuevas variables
- Selección de una muestra de un subconjunto de datos
- Agregación de registros
- Eliminación o sustitución de valores en blanco o ausentes
- Utilización de algún método de imputación

Modalidad de entrega TP N°2:

Una Jupyter Notebook que contenga los gráficos/tablas así como también las explicaciones de lo que fueron realizando, e inclusive las conclusiones o interpretaciones que realicen.

Los criterios de evaluación consideran un tratamiento integral de los conceptos abordados en la materia, adecuada elección y uso de las herramientas y la calidad de los resultados que den cuenta de una adecuada manipulación y tratamiento del dataset. No es necesario que apliquen todos los ítems mencionados o vistos en la materia, con que apliquen algunos métodos bien seleccionados y justificados sería suficiente.

Esto deberán subirlo al aula virtual en el espacio destinado para tal fin. Fecha límite de entrega: **Lunes 17 de Julio.**