

Laboratorio #1 – Ingeniería de Características Parte 1

De forma individual: Para el dataset adjunto, realice las siguientes operaciones de ingeniería de características:

1. Determine que columnas poseen faltantes (NA o Nulos),
2. Determine la proporción de faltantes para cada columna con faltantes, muestre una gráfica de barras con el porcentaje de faltantes para cada columna.
3. Si la variable es numérica realice un análisis de cual sería el mejor método para realizar la imputación de datos, debe implementar un bloque de código donde se muestre cada imputación y su análisis. Recuerde que puede ser:
 - a. Análisis de casos completos (CCA),
 - b. Imputación de media
 - c. Imputación de mediana,
 - d. Generación de números aleatorios en base a la distribución (no visto en clase – opcional).
4. Si la variable es categórica, realice una imputación que considere necesaria, recuerde tomar en cuenta los siguientes aspectos:
 - a. Impacto sobre la variable target.
 - b. Recuerde que el valor que más se repite no necesariamente es el valor que se debe imputar.
 - c. Indicador de clase faltante (Missing indicator).
5. Realice la codificación de variables categóricas, determine cual es le mejor enfoque para cada caso, recuerde que puede realizar los siguientes enfoques:
 - a. One Hote Encodign.
 - b. Codificación por Frecuencia.
 - c. Codificación por Densidad.
6. Identifique cuales variables son discretas y continuas.

Su entregable deberá ser el Notebook con todas sus anotaciones y el dataset en formato CSV con los cambios que determino adecuados según su procedimiento de ingeniería de características.