

Laboratorio #3 – Pipeline - Ingeniería de Características.

En parejas: 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.
3. Si la variable es numérica realice un análisis de cual sería el mejor método para realizar la imputación de datos, debe mostrar un bloque de código donde se muestre cada imputación y su análisis. Recuerde que puede ser:
 - a. Eliminar los faltantes,
 - b. Imputación de media o mediana,
 - c. Valores arbitrarios (extremo derecho y extremo izquierdo),
 - d. Generación de números aleatorios en base a la distribución (no visto en clase).
4. Si la variable es categórica, realice una imputación que considere necesaria.
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. Para todas las variables resultantes realice Feature Scaling según lo considere más pertinente.
7. Crear un modelo de regresión que permita minimizar el RMSE en predicción, para este inciso puede utilizar cualquier modelo de regresión que considere adecuado.

Considerando las instrucciones anteriores deberá crear un pipeline de sklearn para ingeniería de características similar al ejercicio desarrollado en la Sesión de Clase #5.

Su entregable deberá ser el Notebook con todas sus anotaciones y el dataset en formato CSV con los cambios que determino adecuados.