

Actividad 6

Análisis limpieza:

La imputación de datos para el df por medio de métodos como bfill o ffill, generaría valores incorrectos, porque cada registro es único e independiente de observaciones posteriores o anteriores. Las variables numéricas con nulos son categóricas ej. CP, e imputación por medio de mediana o media daría lugar a CP inexistentes. Límite crédito es una variable numérica stored as object y tiene valores str (SD) por lo tanto identifico los str y los reemplazo con la mediana de la columna. Finalmente lo convertí a int. Para no perder registros, lo que hice fue rellenar los nulos de todas las variables con “dato no disponible”. Al identificar que las columnas con nulos eran irrelevantes para el análisis de correlación no lineal porque eran categóricas o numéricas categóricas, decidí únicamente añadir la leyenda y conservar los registros.

Análisis modelos:

Al momento de iterar los modelos, no hubo una sola R2 con un valor < 0.95 . Todos los valores de R2 fueron extremadamente bajos, por lo que ningún modelo tiene un ajuste representativo a los datos reales. Asimismo, al momento de observar las gráficas, es notorio que los valores predichos vs los valores reales no coinciden.