

Análisis Predictivo de Precios Inmobiliarios: Evaluación de Modelos de Regresión y Selección de Variables Clave

Franklin Manuel Manjarres

En este proyecto el análisis se centró en predecir el precio de propiedades inmobiliarias utilizando técnicas de *machine learning*. Tras una exploración inicial, se eliminaron variables redundantes (como `city` y `postal_code`) y se redujo la cardinalidad de `property_type` mediante el análisis de Pareto, conservando solo 9 categorías que representan el 99 % de los datos. Se imputaron valores nulos en variables numéricas (por ejemplo, `size`, `nb_rooms`) con la mediana y se identificaron correlaciones clave: `nb_rooms` mostró la mayor relación con el precio (0.30), mientras que variables geográficas (como `approximate_latitude`) y `size` fueron descartadas por su baja relevancia. Se detectaron *outliers* en precios, con el 2.5 % de propiedades superando 1.275 millones de unidades monetarias, lo que justificó el uso de modelos robustos.

La comparación entre regresión lineal y `HuberRegressor` evidenció que el segundo reduce el error medio absoluto (MAE) de 182,847 a 169,465, mejorando la robustez ante *outliers*. Los coeficientes del modelo Huber destacaron que `property_type_viager` (derecho de uso vitalicio) reduce significativamente el precio, mientras `nb_rooms` y `property_type_propriété` (propiedades exclusivas) tienen mayor impacto positivo. Aunque los errores siguen siendo altos, el estudio subraya la importancia de técnicas de imputación, selección de variables y modelos robustos para tratar con datos desequilibrados o atípicos, manteniendo la regresión como un enfoque base para futuras mejoras.