





Proyecto de Predicción de Precios de Vivienda en California

Problema de Negocio

Este proyecto tiene como objetivo predecir los precios de las viviendas en California utilizando datos geoespaciales y características de las propiedades. Se explora el modelo de machine learning XGBoost para ayudar a comprender cómo varían los precios según la ubicación y las características de las viviendas.

Preguntas Clave

-  **Análisis Inicial:** ¿Qué insights podemos obtener del análisis exploratorio inicial del conjunto de datos?
-  **Transformaciones:** ¿Qué preprocesamiento es necesario para preparar los datos correctamente?
-  **Influencia Geográfica:** ¿Cómo afecta la ubicación (latitud, longitud, distancia a la costa) a los precios de las viviendas?
-  **Características de las Viviendas:** ¿Qué características como el número de habitaciones o el tamaño de las viviendas influyen más en los precios?

Configuración del Ambiente

Asegúrate de tener las siguientes bibliotecas instaladas:




Google Colaboraty

Obtención y Tratamiento de Datos

- **Cargando la Base de Datos:** Los datos provienen del conjunto de datos de California Housing, disponible en formato en el link.

Tratamiento de Datos

Durante el preprocesamiento se realizan las siguientes operaciones:

-  **Manejo de Valores Faltantes:** Se imputa o elimina cualquier valor faltante en las variables clave.
-  **Eliminación de Duplicados:** Se identifican y eliminan filas duplicadas en el conjunto de datos.
-  **Manejo de Outliers:** Se utilizan técnicas como el rango intercuartílico (IQR) para gestionar valores atípicos en el precio de las viviendas.

Normalización de Datos

Se eliminan columnas irrelevantes y se normalizan las variables para asegurar que los modelos tengan un rendimiento óptimo.