

Contexto del proyecto

El mercado inmobiliario es un sector crítico en la economía de cualquier región, donde la valoración de propiedades juega un papel central en la toma de decisiones de compradores, vendedores, desarrolladores y reguladores. Sin embargo, los precios de las viviendas están influenciados por múltiples variables que incluyen características estructurales, ubicación, tamaño, y condiciones del mercado.

Este proyecto busca utilizar el conjunto de datos "**House Prices**", que incluye variables representativas como superficie del terreno, número de habitaciones, calidad de construcción y ubicación. A través del análisis de este dataset, se pretende identificar los factores más influyentes en la determinación del precio de las viviendas y construir un modelo predictivo robusto que permita estimar precios con precisión.

Este dataset:

- Incluye un conjunto de características relevantes para evaluar cómo las propiedades físicas y contextuales influyen en el precio.

Se pide como mínimo:

- a) Realizar un análisis descriptivo de las variables del dataset para comprender su distribución y relación con el precio.
- b) Identificar las correlaciones y evaluar posibles problemas de multicolinealidad entre las variables predictoras.
- c) Ajustar modelos de regresión lineal múltiple, seleccionando el modelo óptimo basado en métricas estadísticas y criterios de simplicidad.
- d) Diagnosticar el rendimiento del modelo mediante análisis de residuos y evaluación de su capacidad predictiva utilizando métricas como R^2 , MAE, RMSE y MAPE.
- e) Interpretar los resultados obtenidos para proporcionar conclusiones sobre el impacto de las características clave en el precio de las viviendas.

Descripción del Dataset

El conjunto de datos **House Prices: Advanced Regression Techniques** incluye 81 variables relacionadas con características de las viviendas y los vecindarios, como:

- **Variables cuantitativas:**

- LotArea: Área del lote en pies cuadrados.
- GrLivArea: Área de la superficie habitable sobre el suelo (en pies cuadrados).
- TotalBsmtSF: Área total del sótano.
- YearBuilt: Año de construcción.
- GarageCars: Número de coches en el garaje.
- ***SalePrice: Precio de venta de la propiedad (variable objetivo).***

- **Variables cualitativas:**

- Neighborhood: Vecindario de la propiedad.
- BldgType: Tipo de vivienda.
- HouseStyle: Estilo de la vivienda.
- ExterQual: Calidad de los materiales y el acabado exterior.
- CentralAir: Si tiene aire acondicionado central (Sí/No).
- MSZoning: Clasificación de zonificación.

Este dataset es ideal porque permite explorar modelos de regresión lineal múltiple y también realizar análisis de covarianza (ANCOVA) para ver cómo influyen diferentes variables de calidad en los precios de las casas.

Rúbrica de evaluación:

Criterio	Descripción	Ponderación (%)
Introducción	Presentación clara del tema, contexto y justificación del proyecto.	10%
Objetivo general	Definición del objetivo principal del proyecto, alineado con el análisis de precios de viviendas.	5%
Objetivos específicos	Establecimiento de objetivos específicos relevantes y medibles para el análisis.	5%
Análisis inicial de los datos	Exploración, limpieza, detección de NA's, análisis descriptivo y visualización preliminar.	15%
Correlación y multicolinealidad	Identificación de correlaciones significativas y evaluación/manejo de multicolinealidad.	15%
Modelado y selección del modelo	Construcción de modelos, comparación mediante AIC y selección del modelo óptimo.	20%
Diagnóstico del modelo	Verificación de supuestos: linealidad, homocedasticidad, normalidad e independencia de residuos.	10%
Evaluación del modelo	Cálculo de métricas como MAE, RMSE, MAPE, R^2 , y validación del modelo si aplica.	10%
Interpretación y conclusiones	Interpretación de resultados, análisis de variables clave y conclusiones con base en los hallazgos.	10%