**EDM**

**task: xai3**

Carlos Barrado Gutierrez, Pablo Torres López, Enrique Ferrer

Contenido

[Introduction 3](#_Toc198231560)

[Exercise 1 3](#_Toc198231561)

[Exercise 2 3](#_Toc198231562)

[Exercise 3 3](#_Toc198231563)

[3.1 Overview and Methodology 3](#_Toc198231564)

[3.2 Impact of sqft\_living: Usable Space Drives Value 4](#_Toc198231565)

[3.3 Impact of bathrooms: A Marker of Comfort and Luxury 4](#_Toc198231566)

[3.4 Impact of floors: Architectural Layout and Price Steps 5](#_Toc198231567)

[3.5 Impact of bedrooms: A Misleading Indicator of Value 6](#_Toc198231568)

[3.6 Model Interpretability and Practical Relevance 6](#_Toc198231569)

[Final Conclusions 7](#_Toc198231570)

# Introduction

-------INTRO-------

# Exercise 1

------EX1--------- PABLO

# Exercise 2

-----EX2---------- QUIQUE

# Exercise 3

## 3.1 Overview and Methodology

Understanding what drives the price of a house is crucial for real estate strategy, both from a business and a client-facing perspective. Using a Random Forest model trained on a representative sample from the kc\_house\_data.csv dataset, we applied Partial Dependence Plots (PDP) to interpret the global behavior of the model regarding four essential features: bedrooms, bathrooms, sqft\_living, and floors. The aim is not only to understand how each variable influences price individually but also to gain actionable insights grounded in the behavior of the model itself.

Imagen que contiene interior, foto, tabla, grande

El contenido generado por IA puede ser incorrecto.

## 3.2 Impact of sqft\_living: Usable Space Drives Value

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

The feature that clearly stood out was sqft\_living. The relationship here is direct and intuitive: as the living area in square feet increases, the predicted price of the house grows almost exponentially. This behavior is reflected both in the steep slope of the PDP and in the model’s internal importance metrics, where sqft\_living shows the highest increase in mean squared error (%IncMSE > 50%). The price curve rises continuously, confirming that this feature holds the greatest predictive power. From a business standpoint, this suggests that highlighting the spaciousness of a property—and even modest extensions—can significantly impact perceived value. The interpretability of this plot is especially strong due to its monotonicity, a desirable property discussed in interpretability theory that eases human understanding.

## 3.3 Impact of bathrooms: A Marker of Comfort and Luxury

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

In contrast, bathrooms also exhibits a generally increasing trend, though with more complexity. For homes with 1 to 4 bathrooms, the effect on price is steady but moderate. However, after 4 bathrooms, the predicted price begins to rise more sharply. This indicates a non-linear effect: having additional bathrooms begins to act as a luxury signal rather than just a functional feature. This reinforces the idea of interaction effects mentioned in the interpretability materials—specifically, how the impact of a variable may change depending on the value range. bathrooms is ranked second in importance, suggesting that high-end buyers are willing to pay a premium for comfort and privacy that comes with additional full bathrooms.

## 3.4 Impact of floors: Architectural Layout and Price Steps

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

The variable floors showed an interesting stepped pattern. The price increases in blocks—most notably from one to two floors and again from two to three—then plateaus beyond that. This suggests that the architectural type (e.g., single-floor vs. duplex vs. triplex) might serve as a proxy for perceived luxury or space distribution, rather than contributing value in a continuous way. While it’s not as influential as other features, the effect is consistent enough to consider it during marketing or design decisions. The plot’s clear stair-step structure reflects the categorical nature of the variable and aligns with the model’s internal logic.

## 3.5 Impact of bedrooms: A Misleading Indicator of Value

Gráfico, Gráfico de líneas

El contenido generado por IA puede ser incorrecto.

Bedrooms, surprisingly, turned out to be the least informative feature of the four. The PDP shows a non-monotonic and somewhat erratic influence on predicted price. In particular, properties with 4 bedrooms seemed to correlate with a slight dip in predicted price compared to 3 or 5-bedroom homes. This counterintuitive effect might arise from confounding factors not captured directly in the model—for example, homes with many bedrooms but smaller living areas might signal outdated layouts or lack of renovation. This is a good illustration of what the course material refers to as the limits of observational interpretation: without considering other latent variables or interactions, the raw effect can be misleading. In terms of strategic takeaway, this tells us that simply increasing the number of bedrooms does not guarantee added value; instead, emphasis should be placed on space quality and design.

## 3.6 Model Interpretability and Practical Relevance

By employing PDPs, which average predictions across all observations, we ensure a global interpretability of the model, giving stakeholders confidence in understanding how these variables drive price regardless of the specific case. This method offers a balance between fidelity and clarity, capturing meaningful relationships while remaining accessible to non-technical decision-makers.

Their main advantage lies in their interpretability and intuitive visualization, making complex model behavior more accessible. However, PDPs assume feature independence and may be misleading in regions with low data density or high feature correlation. For instance, the erratic behavior of bedrooms could be due to hidden interactions not captured in the plot. In the future, combining PDPs with Individual Conditional Expectation (ICE) plots or analyzing interaction effects using H-statistics could provide deeper and more accurate insights.

# Final Conclusions