**EDM**

**task: xai3**

Carlos Barrado Gutierrez, Pablo Torres López, Enrique Ferrer

Contenido

[Introduction 3](#_Toc198304174)

[Exercise 1 3](#_Toc198304175)

[1.1 Partial Dependence Plots and explanations 3](#_Toc198304176)

[Exercise 2 5](#_Toc198304177)

[2.1 PDP 5](#_Toc198304178)

[2.2 Density Distributions 6](#_Toc198304179)

[2.3 PDP Interpretation 7](#_Toc198304180)

[2.4 Density Distributions Interpretation 7](#_Toc198304181)

[Exercise 3 8](#_Toc198304182)

[3.1 Overview and Methodology 8](#_Toc198304183)

[3.2 Impact of sqft\_living: Usable Space Drives Value 8](#_Toc198304184)

[3.3 Impact of bathrooms: A Marker of Comfort and Luxury 9](#_Toc198304185)

[3.4 Impact of floors: Architectural Layout and Price Steps 10](#_Toc198304186)

[3.5 Impact of bedrooms: A Misleading Indicator of Value 10](#_Toc198304187)

[3.6 Model Interpretability and Practical Relevance 11](#_Toc198304188)

[Final Conclusions 12](#_Toc198304189)

# Introduction

Understanding how predictive models make decisions is essential not only for evaluating their accuracy but also for building trust and extracting actionable insights from their behavior. In this report, we apply **Partial Dependence Plots (PDPs)** to a variety of regression models in order to interpret how specific features influence the predicted outcomes. PDPs offer a global perspective on model behavior, allowing us to analyze feature impact while averaging out the influence of other variables. Through this methodology, we examine how environmental and temporal variables affect bike rental demand, as well as how structural characteristics influence housing prices.

The document is structured into three exercises. The first focuses on univariate PDPs for a bike-sharing dataset, assessing how features like temperature, humidity, and wind speed affect predicted rentals. The second exercise delves into bivariate PDPs and feature distributions to uncover interactions between temperature and humidity. Lastly, the third exercise applies the same interpretability technique to a real estate dataset to understand how features like square footage and number of rooms shape housing price predictions. Each section aims to extract practical implications that align with real-world behavior, aiding both technical understanding and business decision-making.

# Exercise 1

## 1.1 Partial Dependence Plots and explanations

We begin our analysis by examining how key features—**days\_since\_2011, temperature, humidity, and wind speed**—affect bike rental predictions in a Random Forest model. The following plots show each variable's marginal impact while controlling for the others, revealing whether their influence is linear, nonlinear, or more complex. This helps us understand how the model captures real-world patterns in the data.

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.Gráfico, Gráfico de dispersión

El contenido generado por IA puede ser incorrecto.

**Days Since 2011 (days\_since\_2011)** The partial dependence plot for days\_since\_2011 shows an upward trend in the predicted number of bike rentals as time progresses since the beginning of the dataset (starting in 2011). This suggests that the model has learned that rentals tend to increase steadily over time. Several external factors might explain this growth, such as urban expansion, improvements in cycling infrastructure, public campaigns promoting bike use, or behavioral shifts among users. The increase is not strictly linear but clearly upward, indicating a consistently positive influence of this variable on rental predictions.

**Temperature (temp)** Temperature has a clearly positive effect on the predicted bike rental counts. The PDP shows that as temperature rises, the model predicts a higher number of rentals. This matches real-world behavior: people are more likely to bike when the weather is pleasant. However, this increase is not linear; the effect becomes stronger at moderate temperatures and then stabilizes or slightly declines at very high temperatures, likely because extreme heat may discourage cycling. Overall, the model has captured a nonlinear but strongly positive relationship between temperature and bike rentals.

**Humidity (hum)** In the case of humidity, the effect on rental predictions is negative. The PDP indicates that as humidity increases, the number of predicted bike rentals decreases. This can be explained by the discomfort of riding in humid conditions, which are often associated with rain or poor weather—conditions that typically discourage bike use. While the negative effect is not as dramatic as with other variables, it is consistent. The model reduces its rental predictions on more humid days, reflecting a moderate negative influence of this variable.

**Wind Speed (windspeed)** Wind speed also shows a negative relationship with bike rentals. The PDP reveals that as wind speed increases, the predicted rental counts decline. This makes practical sense, as strong winds can make cycling physically more difficult, especially for longer trips or for less experienced riders. Although the downward slope in the PDP is not extremely steep, it does show a steady decline, indicating that the model associates windier conditions with lower demand.

**Conclusion** Together, these four variables have different effects on the model’s predictions. Temperature and days since 2011 show a clearly positive influence, reflecting the expected pattern of increasing bike usage over time and during comfortable weather. On the other hand, humidity and wind speed act as deterrents, lowering the model’s rental predictions when their values are high. These results demonstrate that the Random Forest model has successfully learned complex, nonlinear relationships that align well with real-world human behavior regarding bike usage.

# Exercise 2

## 2.1 PDP

Tabla

El contenido generado por IA puede ser incorrecto.

## 2.2 Density Distributions

Gráfico, Histograma

El contenido generado por IA puede ser incorrecto.

## 2.3 PDP Interpretation

The 2D Partial Dependence Plot (PDP) illustrates the joint effect of normalized temperature (temp) and humidity (hum) on the predicted number of bike rentals.

**Temperature effect**

The model clearly predicts higher bike rental counts as temperature increases. This is visually represented by the shift from purple (lower predictions) to yellow (higher predictions) along the x-axis. Warmer days are more favorable for biking.

**Humidity effect**

Humidity has a negative correlation with predicted rentals. As humidity increases (y-axis), the expected bike rentals tend to decrease slightly, especially under moderate temperatures.

**Interaction between features**

The most favorable conditions for bike rentals are high temperatures with low humidity, located in the bottom-right corner of the plot. Conversely, low temperatures and high humidity correspond to the lowest rental predictions, in the top-left corner.

**Conclusion**

The Random Forest model has learned a realistic dependency: people are more likely to rent bikes when it's warm and dry, which aligns with practical human behavior.

## 2.4 Density Distributions Interpretation

**Temperature Distribution**

The normalized temperature shows a bimodal pattern, with two peaks around values 0.35 and 0.7. This suggests that the dataset contains a mix of cool and warm days, but very few extreme cold or hot days.

**Humidity Distribution**

Humidity is skewed to the left, with most values concentrated between 0.4 and 0.7. Extremely dry or very humid days are rare in the dataset.

**Conclusion**

The regions with the highest prediction values in the PDP also correspond to densely populated regions of the data. This means that the model is learning from frequent, meaningful patterns, rather than extrapolating from rare conditions.

# 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

This report has demonstrated the practical value of **Partial Dependence Plots (PDPs)** as a tool for interpreting complex machine learning models in regression tasks. Through their application to both bike rental demand and house price prediction, we have uncovered meaningful relationships between input features and model outputs—many of which align well with real-world expectations.

In the case of bike rentals, variables such as temperature and days since 2011 showed a strong positive influence, while humidity and wind speed had a consistent negative impact. These insights confirm that the model successfully captures seasonal trends and human behavioral patterns. In the housing price model, square footage emerged as the most influential factor, followed by bathrooms and floors, while the number of bedrooms proved less reliable as an indicator of value. These findings highlight the importance of considering not just the quantity but also the quality and context of property attributes.

Overall, the use of PDPs provided a clear, intuitive way to understand global model behavior and supported actionable conclusions from a business perspective. However, we also noted certain limitations—such as the assumption of feature independence and potential misinterpretations in sparse data regions. To address these, future analyses could incorporate complementary tools like Individual Conditional Expectation (ICE) plots or feature interaction metrics to refine the interpretability of predictive models.

By balancing statistical rigor with visual clarity, PDPs prove to be a valuable asset for data-driven decision-making, bridging the gap between technical modeling and strategic insight.