# "XAI: Model-agnostic methods" Report

Carlos Serrano Esteve[1] and Adrián Castillo Portolés[2]

Evaluation & Deployment of Models

## 1 Introduction

The growing complexity of modern machine learning models has driven a parallel demand for transparency and interpretability. Model-agnostic explanation techniques offer a flexible way to probe the behavior of any predictive model without requiring access to its internal structure. In this report, we focus on Partial Dependence Plots (PDPs), a widely used tool that quantifies the marginal effect of one or two input variables on the model's predictions. By applying PDPs to two distinct regression tasks—predicting daily bike rentals and forecasting house prices—we demonstrate how these visualizations reveal actionable insights about temporal trends, weather sensitivity, and key property characteristics, all while treating the underlying Random Forest as a black box.

## 2 Conclusion

Through our experiments, PDPs have proven instrumental in uncovering the relationships that a Random Forest model learns from complex, high-dimensional data. In the bike rental case study, the one-dimensional PDPs highlighted the positive influence of time progression and temperature, as well as the negative impacts of humidity and wind speed on rental volume; the two-dimensional plot further revealed an interaction showing that the benefit of warmer days diminishes under high humidity. In the housing example, PDPs exposed diminishing returns for extra bedrooms and floors, a strong linear price increase with living area, and significant gains per additional bathroom. Overall, Partial Dependence Plots delivered clear, interpretable summaries of model behavior, enabling us to validate domain hypotheses and build trust in the predictions of otherwise opaque Random Forests.

## 3 One dimensional Partial Dependence Plot

In this experiment, we trained a Random Forest model to predict daily bike rental counts (`cnt`) using all available predictors, including temporal, seasonal and weather-related features. To gain insight into how the model uses each feature, we compute one-dimensional Partial Dependence Plots (PDPs), which show the average predicted response as each feature varies over its observed range while all other inputs remain at their original values. PDPs provide an intuitive summary of the marginal effect that a single variable has on the predicted outcome.

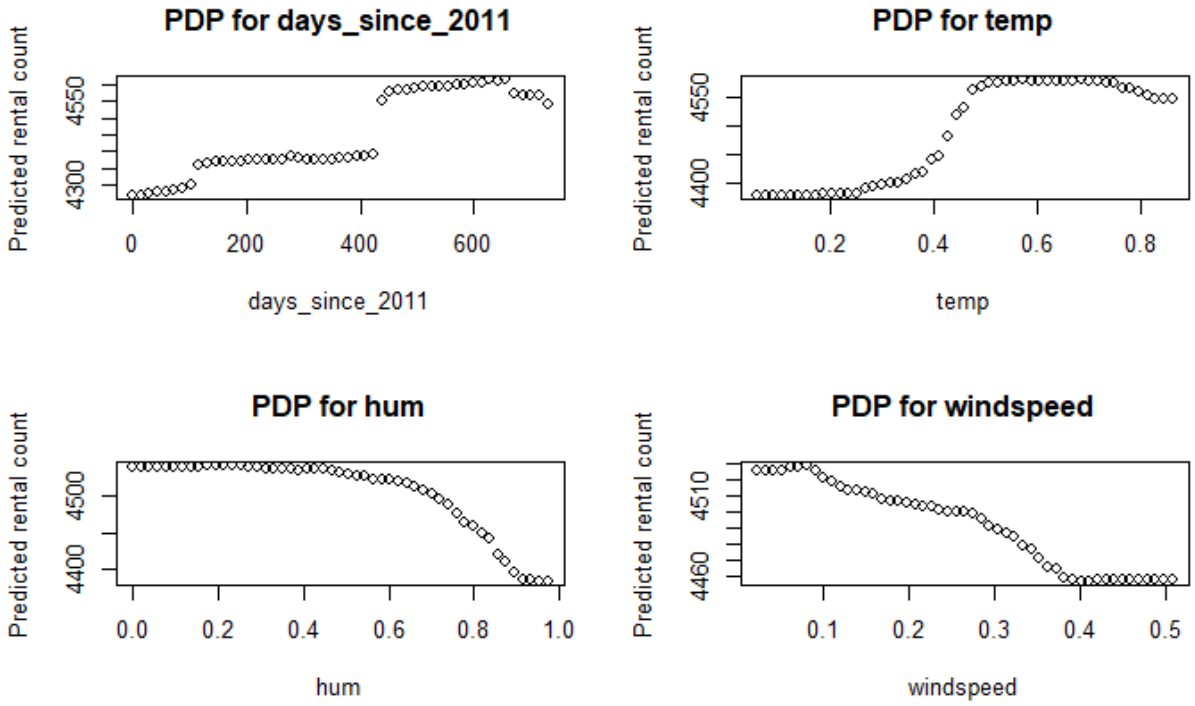


**Fig. 1.** One-dimensional Partial Dependence Plots for `days_since_2011`, `temp`, `hum` and `windspeed`.

The PDP for `days_since_2011` reveals a clear upward shift in predicted rentals around day 550, corresponding to the transition from 2011 to 2012. The model captures an overall increase in demand from approximately 4 320 rentals per day in 2011 to about 4 550 in 2012, consistent with growing usage over time. In the temperature plot, predicted counts rise steadily from cold to mild conditions, climbing from roughly 4 400 up to 4 560 as `temp` increases. Beyond very high temperatures (above 0.8 on the normalized scale), the curve levels off, suggesting that extreme heat yields diminishing returns on rentals.

The humidity PDP shows a nearly linear negative trend: as `hum` increases from 0 to 1, the predicted rental count declines from around 4 520 to 4 400. This indicates that higher moisture—or rainy conditions—discourage bike usage.

Finally, wind speed exerts a modest but consistent downward effect, with gentle breezes (windspeeds near 0.1) associated with about 4 520 predicted rentals, while stronger winds (up to 0.5) reduce expected counts to roughly 4 450. Overall, these plots confirm that milder, dry, and less windy days promote higher bike rental volumes, and they highlight the relative importance of temporal versus weather-related factors in the Random Forest's predictions.

## 4 Two dimensional Partial Dependence Plot

In this section we generate a two-dimensional Partial Dependence Plot (PDP) to explore the joint effect of temperature (`temp`) and humidity (`hum`) on the Random Forest's predicted daily bike rental count. To keep the computation tractable, we begin by drawing a random subset of observations from the full dataset, ensuring that our grid-based evaluation remains efficient without sacrificing representativeness.

Next, we compute the model's average prediction over a fine grid of `temp`–`hum` pairs. All other features in each observation are left at their original values, so the PDP isolates how the two weather variables together

influence the expected `cnt`. The resulting heatmap is constructed with `geom_tile()`, using carefully chosen tile width and height to avoid visual gaps in the surface.

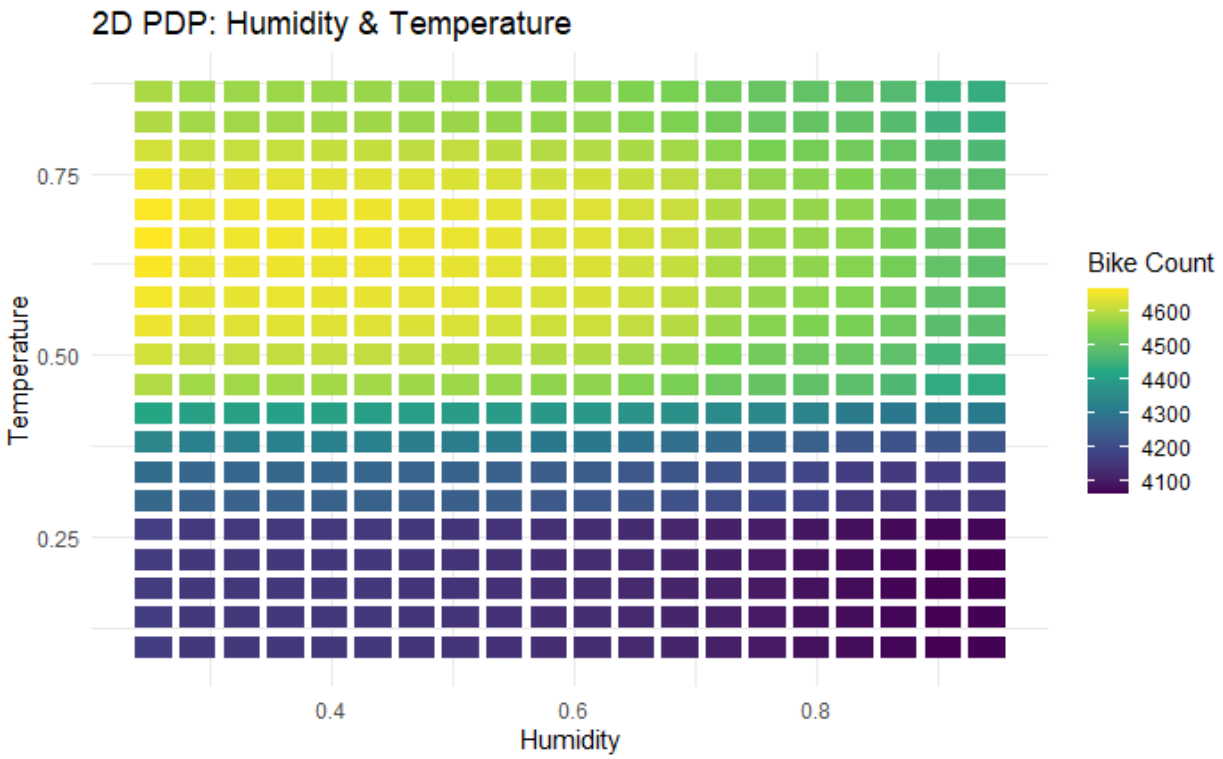


**Fig. 2.** Two-dimensional PDP for `temp` vs. `hum`, with marginal density distributions of each feature.

Figure 2 reveals a pronounced interaction between temperature and humidity. The brightest (yellow) region—indicating the highest predicted rentals—lies at high temperatures combined with low humidity. In contrast, the darkest (blue/purple) area appears where both temperature and humidity are low, corresponding to the lowest rental counts.

Along the top and right margins, we overlay the empirical density distributions of `hum` and `temp`, respectively. These marginal plots highlight that most observations cluster in the mid-ranges (around 0.4–0.7), underscoring that the model's behavior in this central band carries the greatest practical significance.

Overall, the 2D PDP confirms that warm, dry conditions yield the strongest uplift in expected bike rentals, while cool or humid weather dampens demand. Moreover, it shows that the benefit of increasing temperature diminishes once humidity exceeds roughly 0.6, illustrating a subtle but meaningful interaction effect captured by the Random Forest.

## 5 One dimensional Partial Dependence Plot

In this experiment we trained a Random Forest model to predict house prices using the `kc_house_data.csv` dataset. The model was fit on all available features—`bedrooms`, `bathrooms`, `sqft_living`, `sqft_lot`, `floors` and `yr_built`—and then we drew a random subsample of the data to compute one-dimensional Partial Dependence Plots (PDPs). Each PDP shows the model's average predicted price as one feature varies over its observed range, while all other features remain at their original values, thus isolating the marginal effect of that feature.

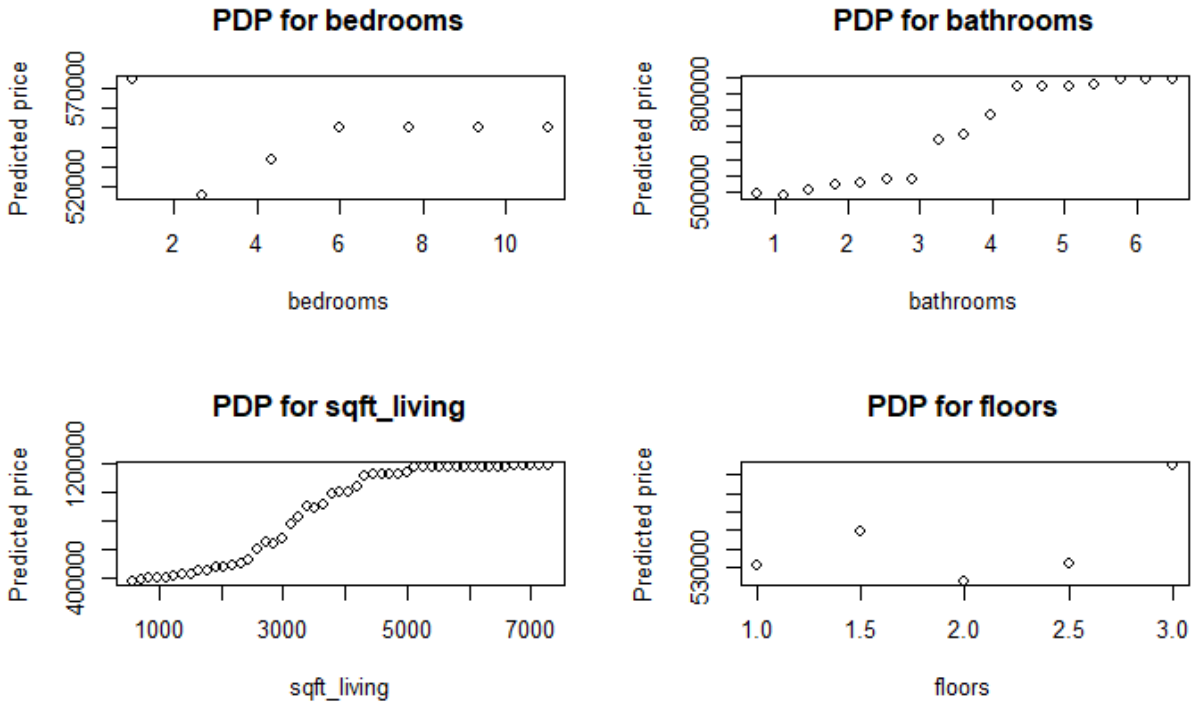


**Fig. 3.** One-dimensional PDPs for `bedrooms`, `bathrooms`, `sqft_living` and `floors`.

The PDP for `bedrooms` indicates a modest upward trend: predicted prices increase as the number of bedrooms rises from two to about six, but beyond six bedrooms the curve flattens, suggesting diminishing returns to adding extremely large numbers of bedrooms. In contrast, the `bathrooms` plot shows a strong positive effect throughout its range—each additional bathroom raises the predicted price markedly until about five bathrooms, after which the impact levels off near the top of the dataset.

Among the continuous predictors, `sqft_living` exhibits the steepest effect: as living area grows, predicted prices climb almost linearly from around $400 000 to over $1 200 000, before tapering off slightly at the very largest sizes. Finally, the PDP for `floors` reveals that single-story homes tend to command lower prices than multi-story properties, with a stepwise jump around two floors and little additional gain for three-story houses. Together, these plots show that living area and bathroom count are the most influential drivers of predicted price in our Random Forest, while bedroom count and number of floors have more moderate but still meaningful effects.

## 6 Conclusion

Through our experiments, PDPs have proven instrumental in uncovering the relationships that a Random Forest model learns from complex, high-dimensional data. In the bike rental case study, the one-dimensional PDPs highlighted the positive influence of time progression and temperature, as well as the negative impacts of humidity and wind speed on rental volume; the two-dimensional plot further revealed an interaction showing that the benefit of warmer days diminishes under high humidity. In the housing example, PDPs exposed diminishing returns for extra bedrooms and floors, a strong linear price increase with living area, and significant gains per additional bathroom. Overall, Partial Dependence Plots delivered clear, interpretable summaries of model behavior, enabling us to validate domain hypotheses and build trust in the predictions of otherwise opaque Random Forests.