Classification and Regression Model Interpretation

Ferran Aragó Ausina, Carles Navarro Esteve, Aleixandre Tarrasó Sorní

Universitat Politècnica de València
{faraaus, cnavest, atarsor}@etsinf.upv.es

Abstract. This report explores model-agnostic interpretability techniques using Partial Dependence Plots (PDP) to analyze how input features influence predictions in two regression tasks. The first part focuses on bike rental demand, using a random forest model trained on weather and temporal variables. One- and two-dimensional PDPs are employed to assess the marginal and joint effects of features like temperature, humidity, and wind speed. The analysis highlights patterns such as increased demand with moderate temperatures and decreased usage under high humidity. The second part examines house price prediction using the `kc_house_data` dataset. PDPs for variables like bedrooms, bathrooms, square footage, and floors reveal nonlinear relationships, diminishing returns, and context-dependent effects, demonstrating the value of PDPs in enhancing transparency and model understanding.

1 Introduction

This report presents a series of analyses focused on model-agnostic interpretability techniques, specifically through the use of *Partial Dependence Plots* (PDP). The main objective is to understand the behavior of machine learning models by isolating and visualizing the marginal effects of input features on their predictions.

The first section investigates the impact of individual variables such as *temperature*, *humidity*, *wind speed*, and *days since the launch of the bike rental system* on the predicted number of daily rentals. A *Random Forest model* is used to estimate demand, and one-dimensional PDPs are generated to interpret the model's behavior. Additionally, a two-dimensional PDP combining *temperature* and *humidity* is included to capture the joint interaction effects on bike rental predictions.

The second section applies PDPs to a different context: the prediction of house prices using the `kc_house_data` dataset. Key predictors such as *bedrooms*, *bathrooms*, *sqft_ living*, and *floors* are analyzed to evaluate their influence on the target variable, using a random forest approximation once again.

Overall, the report aims to demonstrate how PDPs can provide valuable insights into model behavior, enhancing both *transparency* and *interpretability* in regression tasks.

2 Data Preprocessing

For the dataset **Bike-Sharing-Dataset - day.csv** has been performed some transformations in order to prepare the dataset to make the model. First, a *one-hot* encoding was applied to the variable `season`, which originally took values from 1 to 4. To avoid multicollinearity in the model, only three binary variables were included (`season2`, `season3`, and `season4`), using the first one as a reference.

Two additional binary variables were also created from the weather situation (`weathersit`):

- `MISTY`: takes the value 1 if the weather is misty (code 2), and 0 otherwise.
- `RAIN`: takes the value 1 if the weather corresponds to rain, storm, or snow (codes 3 or 4), and 0 otherwise.

Furthermore, three continuous variables were denormalized to return them to their original scales, an important step to make possible interpret the effect of those variables:

- `temp`: rescaled to degrees Celsius using the formula `temp * 47 - 8`.
- `hum`: multiplied by 100 to express it as a percentage.
- `windspeed`: multiplied by 67 to express it in km/h.

In order to incorporate temporal information, the variable `days_since_2011` was created, representing the number of days elapsed since January 1st, 2011, up to the date of each observation.

3 Development of the Project

3.1 Bike rentals

First of all, a Random Forest regression was trained to predict the total number of daily bike rentals (`cnt`) based on features such as *temperature*, *wind speed*, *humidity* and dummy variables (*season.* e.g) The model was trained using the `randomForest` package with 100 decision trees and feature importance computation enabled.

Before analyzing the graphs, let's understand what a Partial Dependency Plot is. These graphs show how the value of the target variable varies with respect to the value of the selected variables. To do this, the value of the variable to be explained changes, assigning values throughout the variable's range, keeping the rest constant for each sample, and obtaining the prediction in each case. The average is then calculated to obtain the overall behavior of the variable.

To interpret the model's predictions, Partial Dependence Plots (PDPs) were generated to analyze the effect of individual variables on the predicted number of rentals. These plots provide a clear understanding of how each feature influences the target variable while averaging out the impact of all other inputs.

The marginal effects of the four most relevant continuous features were visualized: *temperature*, *humidity*, *wind speed*, and *days since 2011*. Each plot includes a rug plot at the bottom, which shows the density distribution of the corresponding variable in the dataset. These marks help identify which input values are more common and give context to the model's learned behavior.

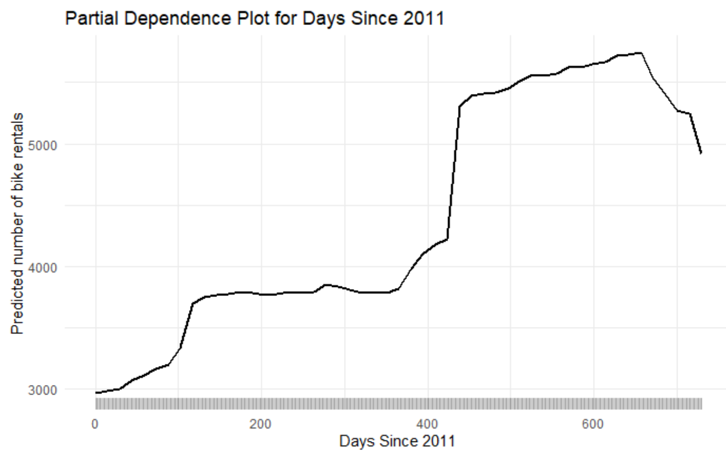

Fig. 1. Partial Dependence Plot for *days since 2011*.

The PDP for *days since 2011* **Fig. 1** reveals a steady increase in predicted demand as time progresses. This trend likely reflects the growing popularity and adoption of the bike-sharing system. Around day 400, a noticeable jump in predicted rentals occurs, followed by a plateau at higher demand levels, suggesting system consolidation and user loyalty over time. As expected in a time-indexed dataset, the rug plot shows a uniform distribution of observations, since each day is represented once.

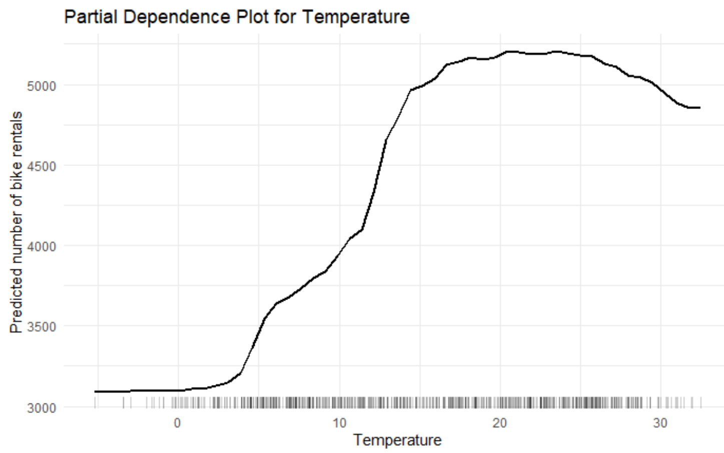

Fig. 2. Partial Dependence Plot for *temperature.*

The plot for *temperature* **Fig. 2** shows a strong positive effect on bike rentals. Demand increases rapidly from 5°C to about 25°C, after which it stabilizes and slightly decreases at extreme temperatures. The rug plot confirms that most temperature observations fall within this optimal range. These results suggest that moderate temperatures create optimal conditions for biking.

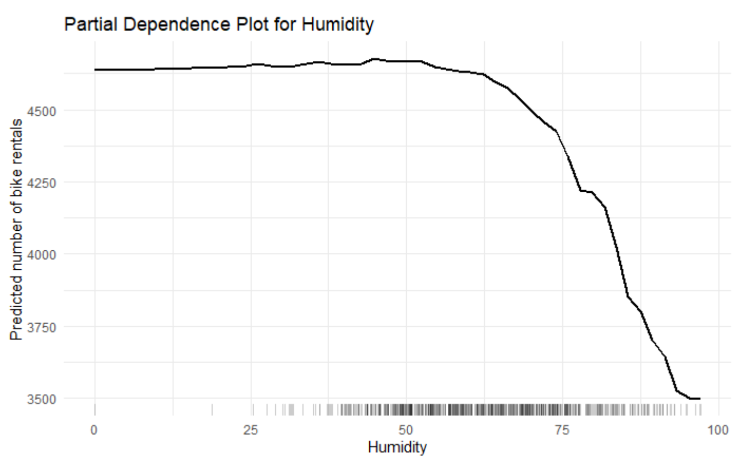

Fig. 3. Partial Dependence Plot for *humidity.*

In the case of *humidity* **Fig. 3**, the predicted number of rentals remains stable up to around 60–70%, after which it drops significantly. This suggests that high humidity discourages bike usage, possibly due to discomfort or perceived inconvenience. At very low humidity levels (below 40%), the predicted demand also remains flat; however, the rug plot reveals an almost complete absence of observations in this range. As a result, the model is likely extrapolating based on insufficient information, and this part of the curve should be interpreted with caution. Conversely, while the decline beyond 70% appears steep, this region is also underrepresented in the data, which may affect the robustness of the learned relationship. Most data points are concentrated between 40% and 60%, supporting the stability observed in that interval.

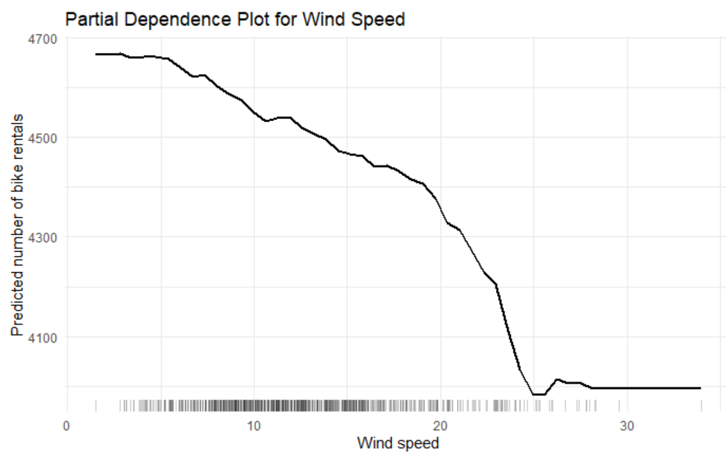

Fig. 4. Partial Dependence Plot for *wind speed*.

In the case of *wind speed*, the Partial Dependence Plot **Fig. 4** shows that the predicted number of rentals remains relatively stable across low to moderate values, with a slight downward trend starting around 10–15 km/h. Most observations are concentrated between 5 and 15 km/h, as indicated by the rug plot, which suggests that the model has sufficient data to learn a meaningful pattern in this range. As wind speed increases beyond this point, the prediction continues to decrease gradually, indicating a potential discouraging effect of stronger winds on bike usage. However, since very few samples exist above 20 km/h, the model's output in that region should be interpreted with caution. Overall, the effect of wind speed is subtle but consistent with the expectation that adverse weather conditions can negatively impact demand.

After having carried out the analysis of the effects of individual variables on the predictions, we now turn to the analysis of the interaction between two of them: temperature and humidity. To represent this interaction, we use a Partial Dependence Plot (PDP) with two variables — *humidity* (x-axis) and *temperature* (y-axis) — showing their joint effect on the predicted count through a color gradient. This visualization helps us understand how changes in *humidity* and *temperature* affect the model's predicted outcome. The color scale represents the predicted value, with redder areas indicating higher predicted counts and bluer areas indicating lower ones.

Red regions, located mostly in the upper center of the plot, correspond to higher predicted counts, around 5000. In contrast, the blue areas in the lower right corner indicate lower predicted counts, around 2500.

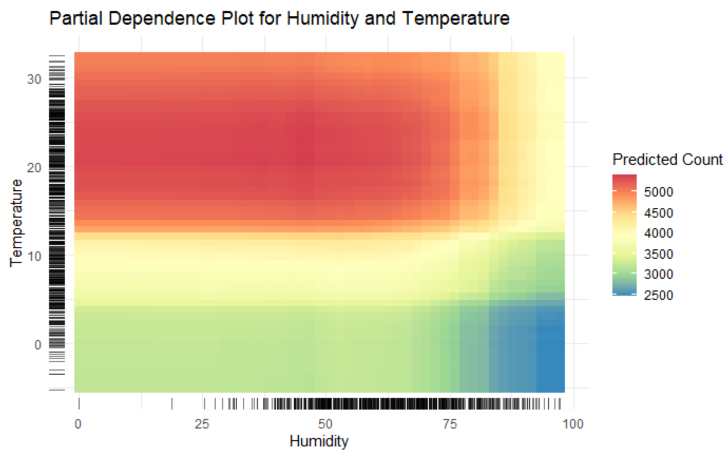

Fig. 5. Partial Dependence Plot for *temp+hum*.

From the plot **Fig. 5**, several conclusions can be drawn. Regarding the effect of *temperature*, higher values — above approximately 10°C — are associated with increased predicted counts that remain consistently until a *humidity* of more than 75-80 % of the humidity level. This is illustrated by the large red area at the top of the plot. On the other hand, at lower temperatures, the predicted number of bikes decreases substantially, especially under high humidity conditions, where the plot shows that in that cases the predictions of rent bikes have the lowest values.

As for *humidity*, its effect is more pronounced at low temperatures. In this case, increasing humidity leads to a sharp drop in predicted counts, as shown by the transition from green to blue in the lower-right part of the plot. At higher temperatures, humidity has a noticeably smaller effect on the predictions.

Additionally, the tick marks along the axes represent the distribution of the data used for training. These indicate that most data points correspond to moderate values of humidity and temperature. As a result, predictions in regions with extreme values (either very low or very high) should be interpreted with caution, as the model had limited data to learn from in those areas.

3.2 House price

In this case, we want to predict the price of a house based on several attributes. These attributes include the number of bedrooms, bathrooms, and floors, as well as the year of construction, living area, and lot size. For the prediction, we use a *random forest model*.

Regarding the model, we want to analyze the effect of the number of bedrooms, bathrooms and floors on the price. To do this, we use a again a **Partial Dependence Plot** for each variable so we can analyze how each one behaves in relation to the target variable.

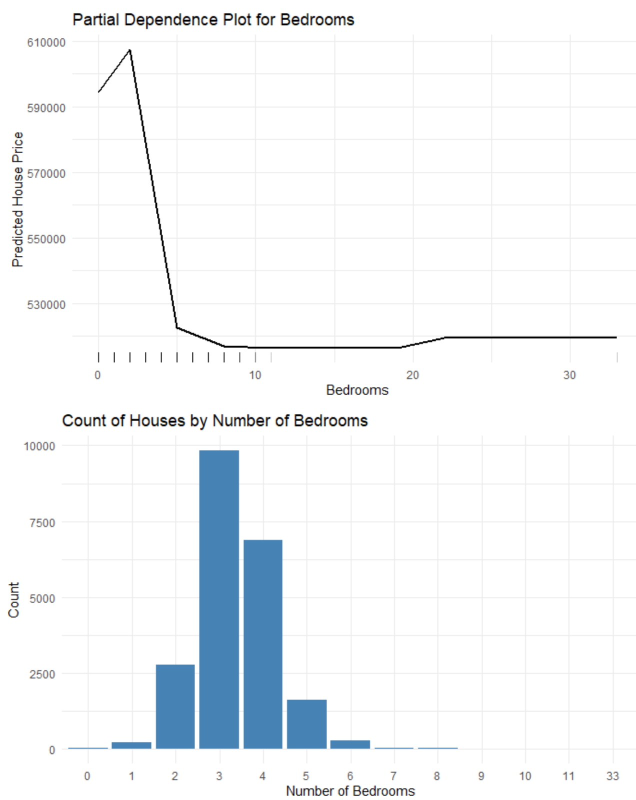

Fig. 6. Influence of Bedrooms on the price

Starting with the number of bedrooms, whose PDP is shown in **Fig. 6**, it is important to note that this is a discrete variable. As such, a continuous distribution will not appear in the plot; instead, all values are concentrated at integer levels. The plot shows that values below 10 exhibit more intensity, indicating that most observations fall within this range. To better illustrate this distribution, we created an additional plot that displays the count of instances for each number of bedrooms, where we can see that homes with 2 to 5 bedrooms are the most common in the dataset. Therefore, the conclusions we draw can only be based on the graph's behavior with these values. This is because the PDP performs this process with all the values in the range, but the model hasn't actually learned with all the values, but rather from the values that actually appear in the dataset, so incorrect conclusions can be drawn.

From what can be seen in these values, it seems to intuit that initially, increasing the number of bedrooms increases the price of the home, but beyond three bedrooms, this value decreases, making it seem as if an excess of bedrooms causes buyers to lose interest.

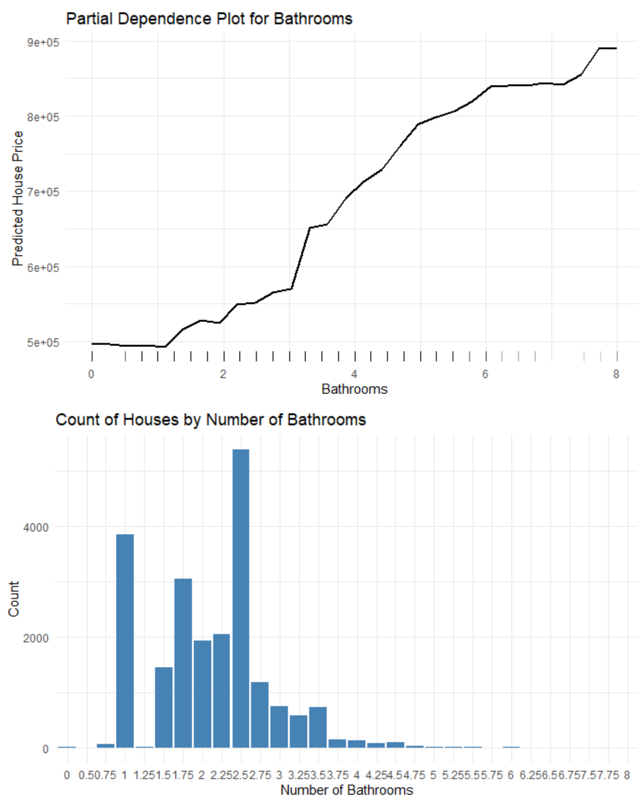

Fig. 7. Influence of Bathrooms on the price

In the bathrooms plot **Fig. 7**, by examining the distribution histogram makes it clear that fully half of all homes cluster between one and two bathrooms, with a pronounced mode at two baths and secondary peaks around one-and-a-half and two-and-a-half baths. Single-bath properties number in the low thousands, three-bath homes only in the high hundreds, and entries below one or above four baths are exceedingly rare. Consequently, our model's behavior between roughly one and three bathrooms is well supported by data, whereas its extrapolations below a half-bath or beyond four full baths rest on very few examples and merit caution.

Within that well-populated one-to-three-bath range, the PDP curve begins nearly flat—adding a half or single bath to a very small home yields only a modest price uplift. As the count climbs toward three baths, the slope steepens noticeably, indicating that buyers prize the additional convenience and en-suite privacy that a third full bathroom brings. Beyond about three and a half baths, however, the curve turns over and flattens: in the luxury segment, each extra bath delivers progressively smaller marginal gains in predicted price, suggesting that high-end finishes, lot characteristics, and overall design quality become more decisive than sheer bath count.

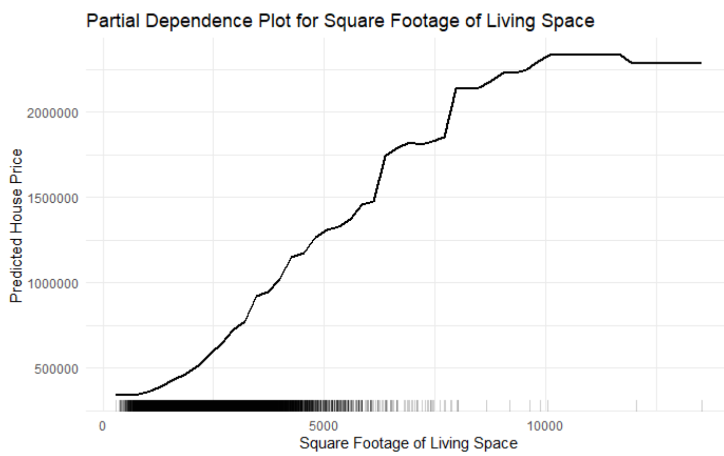

Fig. 8. Influence of Squared Footage on the price

The **Fig. 8**, shows the PDP for the Squared Footage variable. Across most of the observed range—particularly from the smallest homes up through roughly the upper-mid sizes—the density of tick marks shows that the model is well supported by ample training data. In these zones, our interpretations of how square footage influences price are grounded in frequent examples, whereas the sparsely populated extremes (very compact or exceptionally large properties) should be treated with more caution, since the model relies on fewer reference points there.

Within that well-populated middle band, the model predicts that each additional square foot adds increasing value: modest gains at the low end transition into steeper price rises as homes grow from entry-level to larger family proportions. This suggests that buyers reward the jump from basic layouts into more spacious configurations—likely reflecting the premium on extra bedrooms, larger living areas, or open-plan extensions that become possible in this segment.

Beyond this central zone, however, the curve flattens. Once a property exceeds what might be considered a generously sized family home, further increases in living area yield progressively smaller price increments. In practice, this tapering indicates diminishing marginal returns on square footage alone, as other factors—such as lot characteristics, design quality, or local market ceilings—begin to dominate valuation for very large residences.

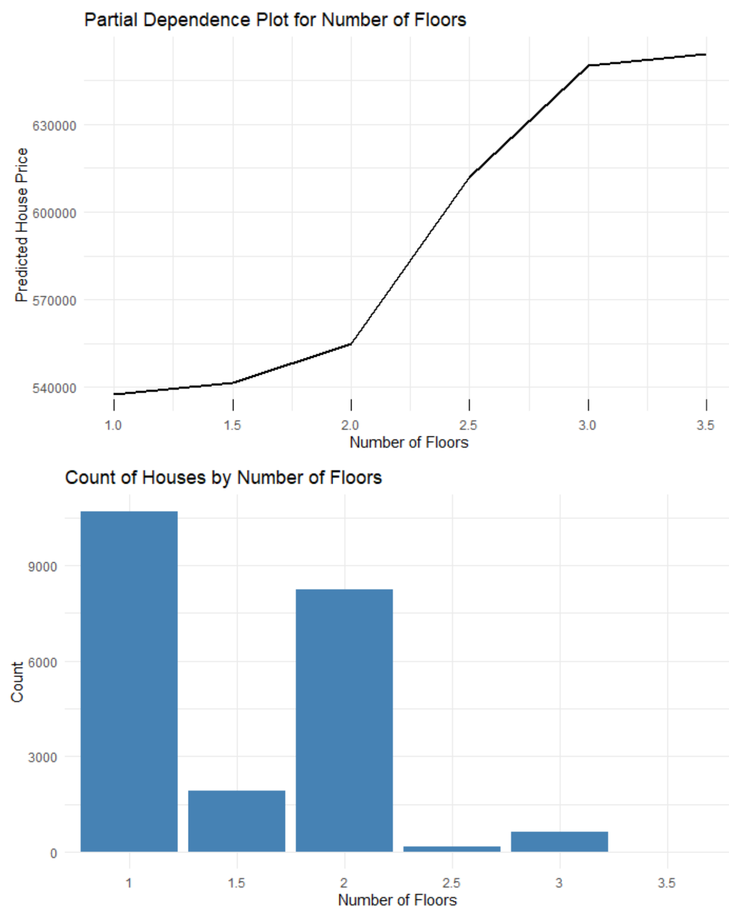

Fig. 9. Influence of Floors on the price

The floors plot, **Fig. 9**, shows that the vast majority of properties are one- or two-story homes—approximately ten thousand single-story and eight thousand two-story houses—followed by a couple of thousand split-levels (1.5 stories).

Three-story residences are scarce (a few hundred), and anything beyond that is almost nonexistent. Thus, the PDP from one to two stories is underpinned by abundant data, while the curve's shape at three or more floors depends on less observations.

Between one and two floors, the model predicts only a modest price increase—reflecting a typical buyer's slight premium for going from a ranch-style to a two-level layout. However, as homes extend into two-and-a-half and full three-story configurations, the PDP slope grows steeper, signaling that added vertical living space commands a stronger premium when it meaningfully expands usable area without increasing lot size. Once past three stories, the curve levels off again, indicating diminishing returns per additional floor as practical factors—such as accessibility, building complexity, and local market preferences—begin to outweigh the simple story count.

4 Conclusions

Throughout this report, Partial Dependence Plots(PDPs) were applied to Random Forest models in two distinct contexts: the prediction of bike rentals and house prices. These plots enabled the interpretation of complex, non-linear models by isolating the marginal effects of individual variables on the target.

In the bike rental case, PDPs revealed that temperature and days since the system's launch were strong positive drivers of demand. Moderate temperatures (around 20–25°C) corresponded to the highest predicted rental counts, while service maturity over time showed a clear trend of increasing usage. Conversely, high humidity and strong wind conditions had a negative impact on predictions, though their effects were mostly limited to underrepresented regions of the dataset. The 2D PDP further confirmed that favorable conditions—warm and moderately dry weather—maximize demand, while colder, humid days strongly reduce it.

For the house pricing model, based on the analysis of Partial Dependence Plots (PDPs) for the variables *bedrooms*, *bathrooms*, *square footage*, and *floors*, it is possible to observe that the model captures meaningful and interpretable relationships between these features and house prices, particularly in regions where data is densely concentrated. The PDPs for bathrooms and square footage show a clear pattern of increasing price with higher values, followed by a plateau or diminishing returns in the higher ranges. This reflects typical market dynamics where moderate improvements bring noticeable value, but excessive quantities (e.g., too many bathrooms or overly large homes) contribute less significantly to price. Similarly, the PDP for floors indicates a premium for additional levels, especially between one and three stories, though the effect also tapers off beyond that point.

However, caution is necessary when interpreting model behavior in sparsely populated regions of the feature space. For instance, properties with very high numbers of bedrooms or bathrooms are underrepresented in the dataset, which makes the corresponding PDP segments less reliable. The combination of PDPs and distribution plots effectively highlights this limitation and emphasizes the importance of grounding interpretation within well-supported data ranges. Overall, the results demonstrate the utility of PDPs in revealing nonlinear and context-dependent relationships, helping to enhance model transparency and supporting more informed decision-making in real estate pricing.