XAI 3: MODEL-AGNOSTIC METHODS

Pablo Espinosa Martínez

Juan Francisco Pallardó Latorre

Alejandro Hidalgo Raga

INDEX

1. Introduction

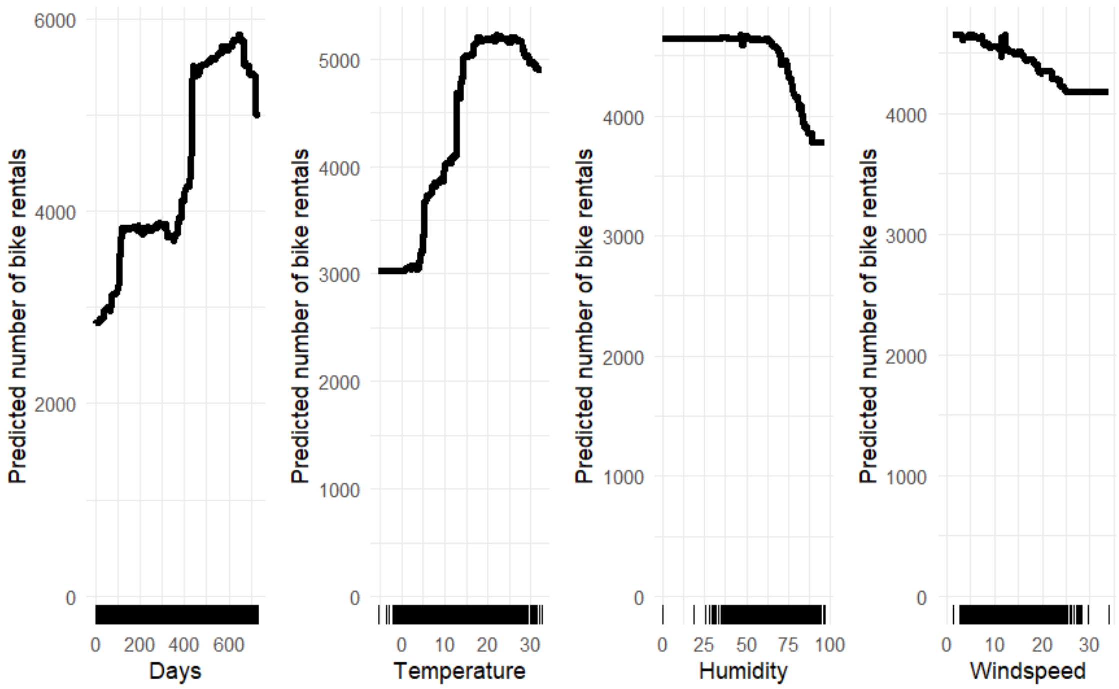
Partial Dependence Plots (PDP) can be very useful and interesting to show the dependence between the target response and a set of input features of interest. Intuitively, we can interpret the partial dependence as the expected target response as a function of the input features of interest.

With this model-agnostic method, we can acknowledge the influence of any predictive variable on the output feature using any model. In this case, we are going to use a random forest for our analysis, but PDPs are compatible with every model.

We will use the bike rental dataset seen in previous exercises to build a one-dimensional partial dependence plot and a bidimensional partial dependence plot. Additionally, we will use a dataset about house prices to predict the price of a house from features such as the number of bedrooms and bathrooms. We will also build a one-dimensional PDP in this case. This dataset is extracted from the sales in King County, Washington State, USA, from May 2014 to May 2015.

2. One dimensional Partial Dependence Plot

Firstly, we are going to apply a PDP to a random forest using the bike rental dataset. With this model, we will be predicting the number of bikes from other features. The features used to predict the response variable are the same as we introduced in previous practical exercises.



Firstly, as we can see, the number of bikes increases over time. However, from the summer of the first year until the end of the year (from 150 to 360 days since 2011), the value remains constant and even decreases at the end. In the second year, the number of bikes grows really fast. However, in the fall of the second year, it decreases again.

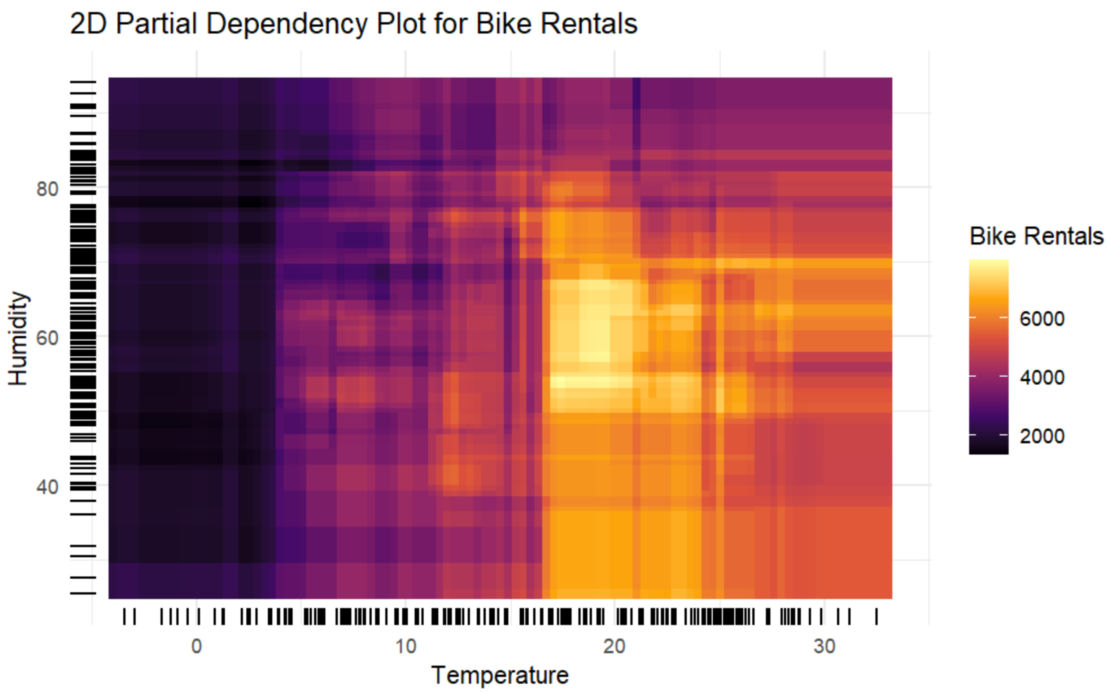
Secondly, from the temperature, when the temperature is between 0 and 8 degrees, the number of bikes remains constant in 3000 bikes. After 8 degrees, in general terms, as the temperature increases, the number of bikes rented also increases. However, when it gets around 19ºC the number of bikes remains constant again. Finally, when the temperature is higher than 30ºC, the number of bikes drops slightly.

Thirdly, the humidity remains constant before a percentage of humidity of 60%. When the percentage increases more than 60%, the number of bikes rented goes down. However, we have to be careful with the conclusions extracted from the 0% of humidity and 25% of humidity, as we have less samples.

Finally, as the windspeed increases, the number of bikes decreases. However, when the windspeed is higher than 25, the number of bikes remains constant. As the previous example, we have just a few samples when the windspeed is higher than around 27, so must be cautious.

3. Bidimensional Partial Dependency Plot

Now we are going to show a PDP with 2 dimensions from a random forest. In the model, we try to predict bike rentals from two features, temperature and humidity.



In the axis we can see the features represented, while the number of bike rentals is indicated by the color, in a scale from around 2000 to 8000. The small black marks along the axes represent individual values of temperature and humidity in the data used to generate the plot. These marks indicate the density of the data in those regions of the feature space.

When interpreting this plot, it's important to focus on the different regions of the plot and the difference in the gradient of colors. It is also worth noting that the density is key to take important conclusions, because as more samples we have, the results will be more accurate.

In the left part of the plot we can see the darkest color for the bike rental when the temperature is lower than 5ºC, independent of the value of humidity. We can extract the influence of low temperatures on less bikes rented.

When the temperature is cold but not freezing, between 5ºC and 15ºC, it is obvious a change in the color, as more bikes are rented. Focusing on the interaction with the other variable, now lower values of humidity mean more bikes rented.

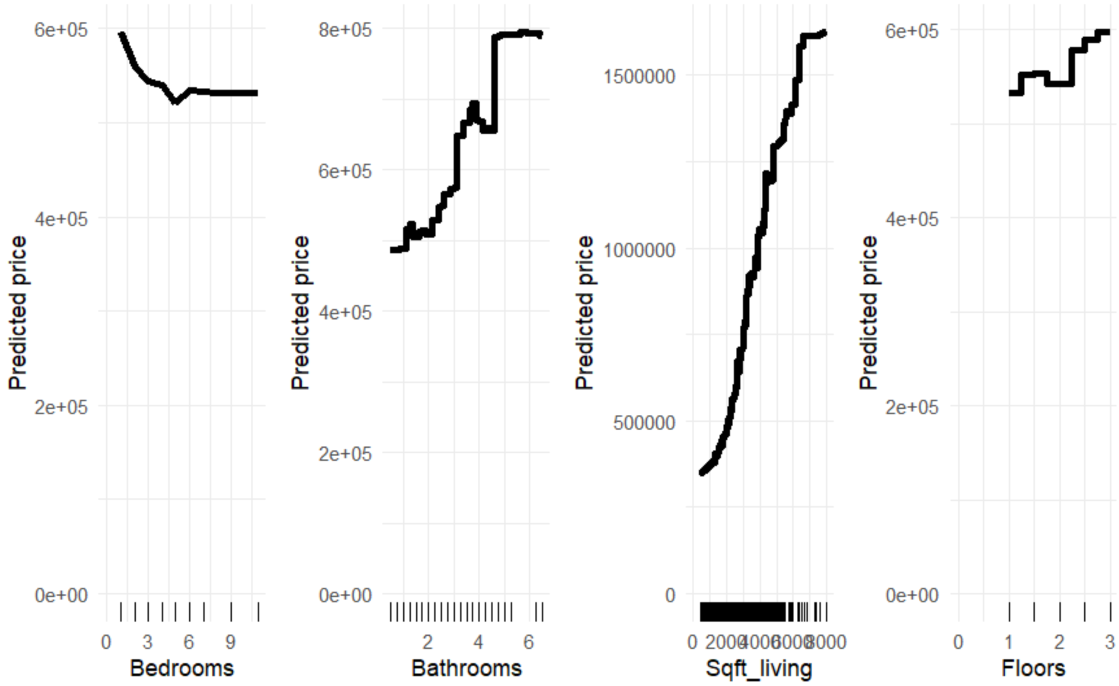
For moderate temperatures, between 15ºC and 25ºC it is really notable the change on bikes rented, arriving to the maximum point if these conditions are given. But it does not only depend entirely on the temperature, the impact of the humidity is also clear, as with high and low values of humidity the bike rentals decrease.

Finally, with high temperatures the bike rentals value is not that high but it is still higher than during cold days. Again, we can see how high values of humidity mean less bikes rented.

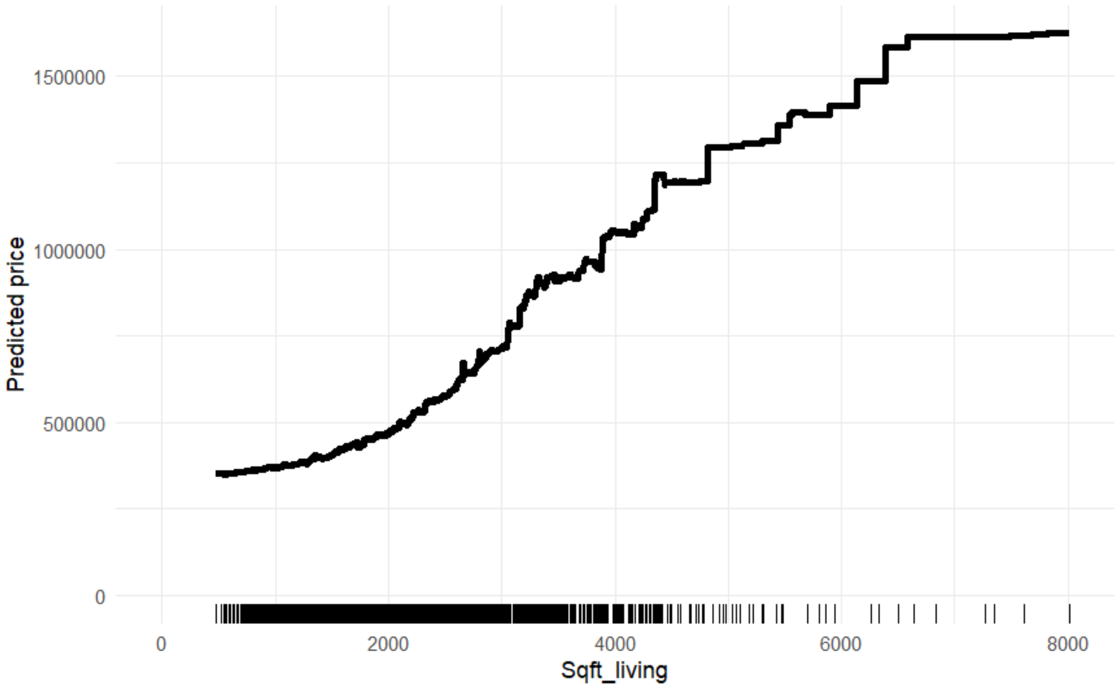
To conclude, the plot shows how both features interact and how they affect to the response variable, resulting in the conclusion that bike rentals are most frequent in conditions of moderate temperature (around 20 degrees Celsius) and moderate humidity (40-60%). Extremes, either in low temperature with high humidity or high temperature with high humidity, tend to show a decrease in the number of bike rentals.

4. PDP to explain the price of a house

As we have done previously, we will use the PDP to try to explain the house price prediction made by a random forest model. The random forest model has been trained with the features bedrooms (number of rooms), bathrooms (number of bathrooms), sqft_living (square feet of living space), sqft_lot (square feet of the lot), floors (number of floors), and yr_built (year built). Now we will look at the PDPs for the variables sqft_living, bedrooms, bathrooms, and floors, and we will try to explain the results:

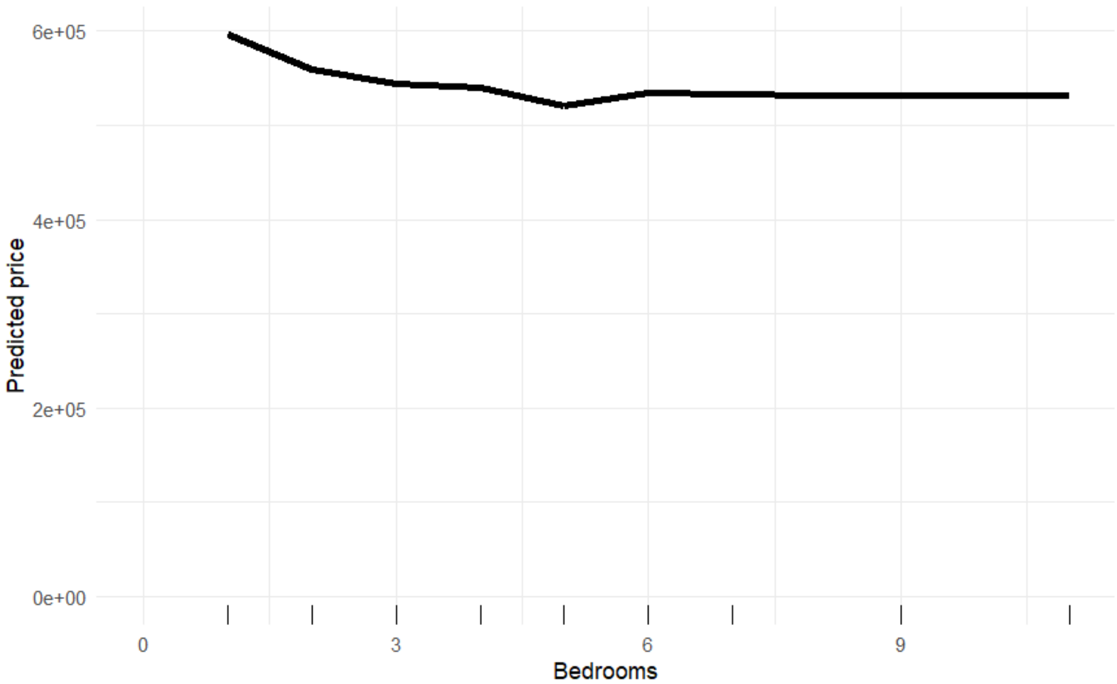


For these types of plots, it is important to observe the data distribution, which is visible at the top of the X-axis. It is important to observe the distribution because the prediction will be more reliable where there is a higher concentration of points (darker). First, we will focus on sqft_living feature:



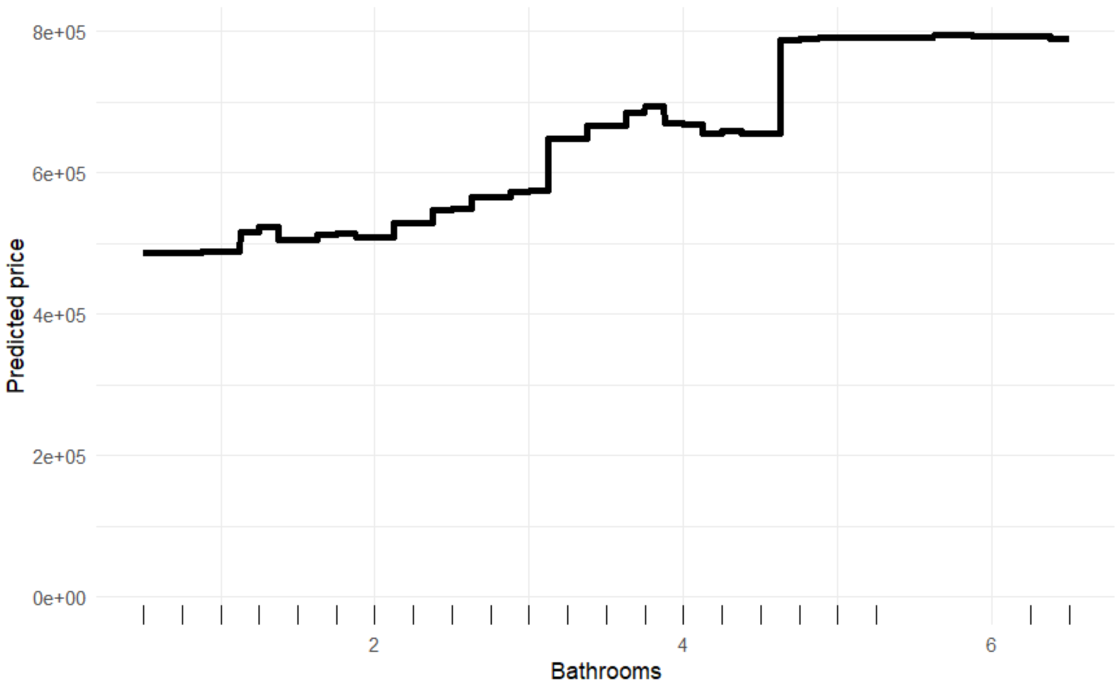
As can be seen in the graph, as the square footage of living space increases, the price of the house goes up. This makes sense and is common sense. What is noteworthy about this graph is that the prediction of this model is very reliable from 480 square feet to 5000 square feet. Beyond 5000 square feet, the distribution decreases, making the prediction less reliable, which in some cases leads to the observation that despite the increase in square footage of living space, the price remains constant.

Now, we must move on the bedrooms feature:



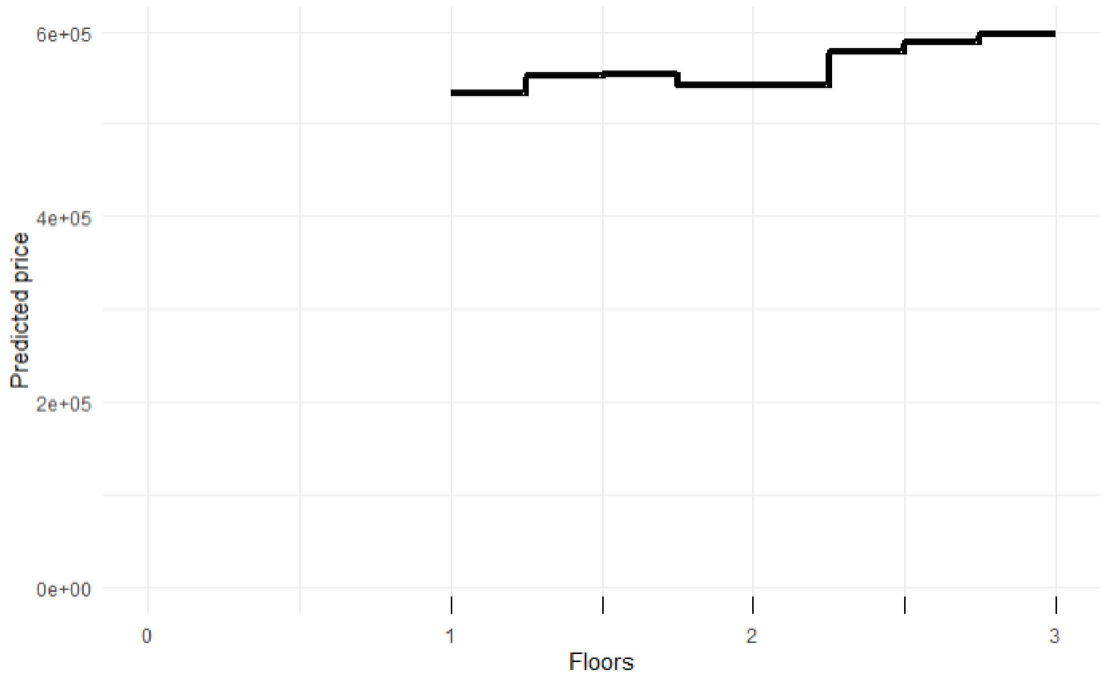
We can see a predicted price decrease between 1 and 3 bedrooms. Despite that, after this point predicted price stabilizes and remains constant, regardless of the increase in the number of bedrooms. We can consider the prediction reliable from 1 to approximately 7 bedrooms. Beyond 7 bedrooms, the distribution becomes weaker, thereby reducing the reliability of the prediction. The observed trend makes practical sense as well. Most typical family homes have between 2 to 4 bedrooms, and beyond this range, additional bedrooms may not contribute significantly to the overall value of the home.

Then we will focus on bathrooms:



As can be seen, between 0 and 2 bathrooms there is a slight increase in the predicted price. This increase is enhanced in the range of 2 to 4 bathrooms, reaching a larger increment around 5 bathrooms. However, beyond 5 bathrooms, we find that the price stabilizes, but we cannot draw significant conclusions due to the distribution. This trend makes sense as more bathrooms mean more luxurious houses, raising the price.

Lastly, we'll talk about floors feature:



Thanks to the distribution, we can say that between 1 and 3 floors the predicted price increases slightly, with the majority of the price increase occurring between 2 and 3 floors. This makes sense and is related to the variable sqft_living, as more floors mean more square feet of living space, which translates to an increase in the house price.

5. Conclusion

In conclusion, along the document there have been evidences about the usability of Partial Dependence Plots that have brought to light how important they could be for the interpretability of complex machine learning models. Apart from this, we have learned that they can be used with one or two dimensions indistinctively, showing how the value of a feature affects the prediction or how the interaction between 2 concrete values of 2 variables contribute to the output. This helps everyone involved to trust the model's results and make better decisions based on them. By making complex models easier to understand, PDPs help ensure that the models are used correctly and effectively. However, when we increase the number of variables, this plot becomes impossible to represent.