XAI3: Model-agnostic methods

Introduction

In the following work we present an analysis of random forest models using a model agnostic explainability method: Partial Dependence Plots (PDP). As a model agnostic method PDP can be utilized for different kinds of models as it does not exploit any model specific features. PDPs visualize the relationship between input features and the model predictions. PDP are computed by fixing the investigated feature in its range and averaging the resulting predictions. This gives a plot of average prediction values against the values of the investigated variable. They help to understand how the model's predictions change with respect to the specific features being analyzed, providing insights into feature importance and effect. Bidimensional PDPs can be used to investigate the interaction effects.

Partial Dependence Plot for Bike Rental Prediction

We used PDPs to investigate the influence of *days since 2011*, *temperature*, *humidity* and *wind speed* on the predicted bike rentals.

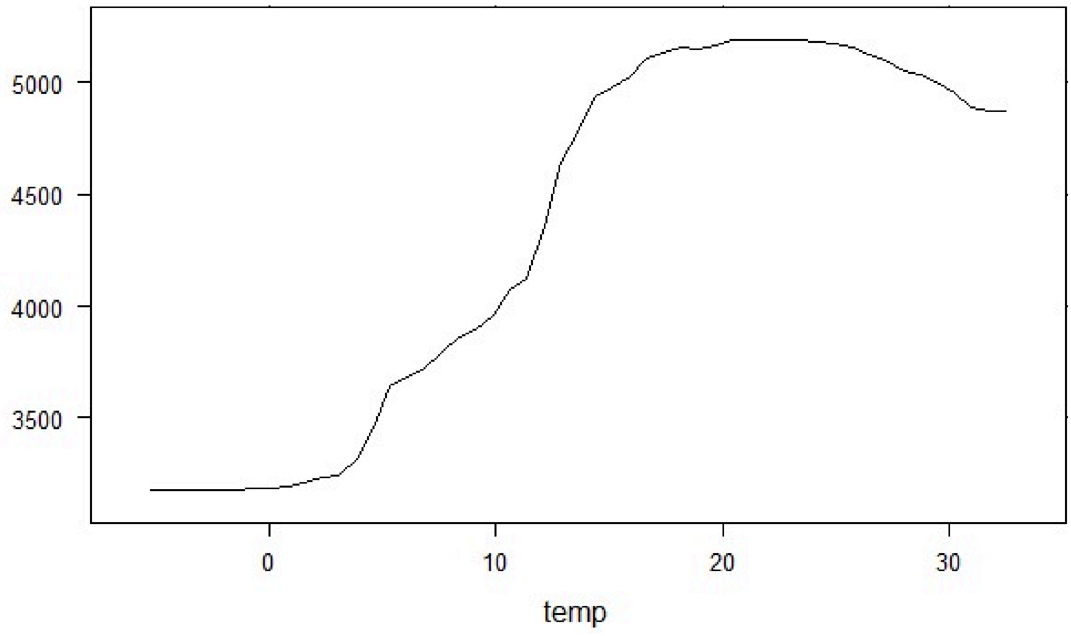


Figure 1: PDP for Temperature

The first figure (Figure 1) shows the change in average number of predicted bike rentals in respect to different temperature conditions. We observe that bike rental numbers are predicted to be lowest when temperatures are low and rise with higher temperatures. They reach their maximum of over five thousand rentals around 22 degrees. After surpassing around 25 degrees the predicted number of rented bikes drops again indicating that high temperatures again lead to the prediction of less bike rentals.

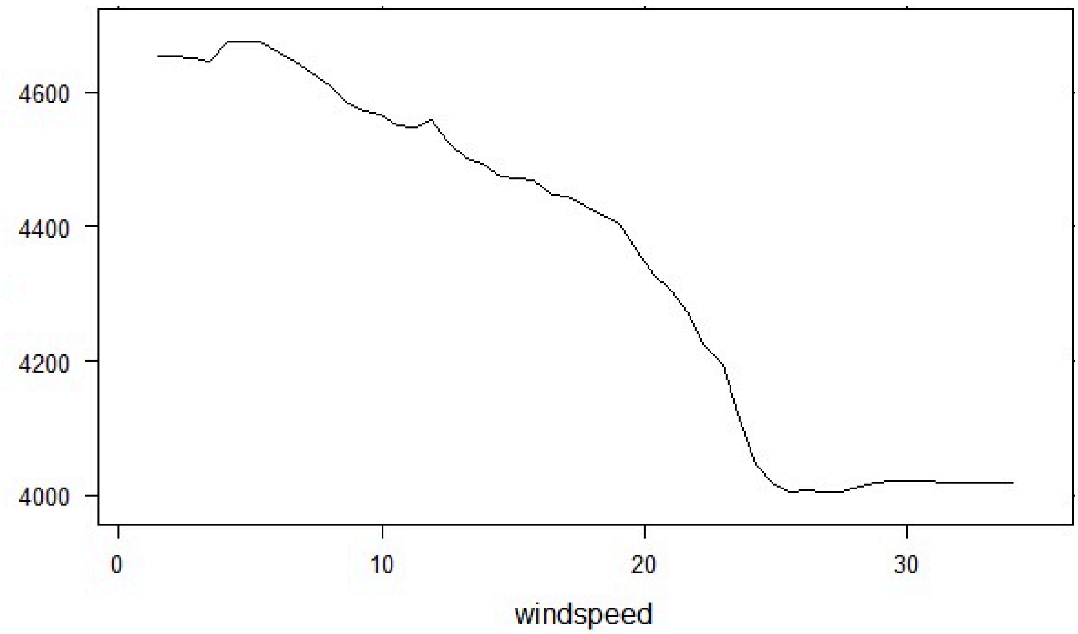


Figure 2: PDP for wind speed

Figure 2 shows the change in average number of predicted bike rentals in respect to wind speed. We observe that bike rental numbers are predicted to be highest with low wind speeds and reduce with higher wind speeds. The number of predicted rentals starts decreasing around 5 kph wind speed. After surpassing around 25 kph the predicted number of rented bikes reaches a relatively stable minimum.

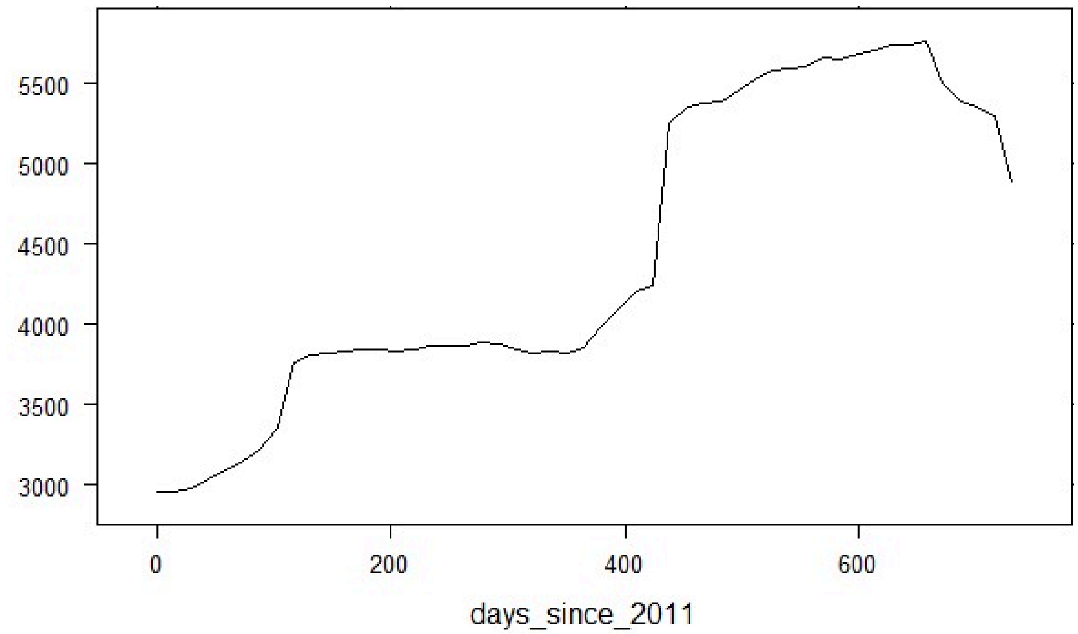


Figure 3: PDP for days since 2011

The PDP for days since 2011 (Figure 3) shows the change in average number of predicted bike rentals in respect to the number of days since 2011. We observe that the number of predicted bike rentals rises over the years. Interestingly this increase is not linear or following another obvious pattern. It appears like usage increases in waves. They reach their maximum of over 5500 rentals around 650 days after 2011. This is followed by an unexpected decrease in predicted rentals.

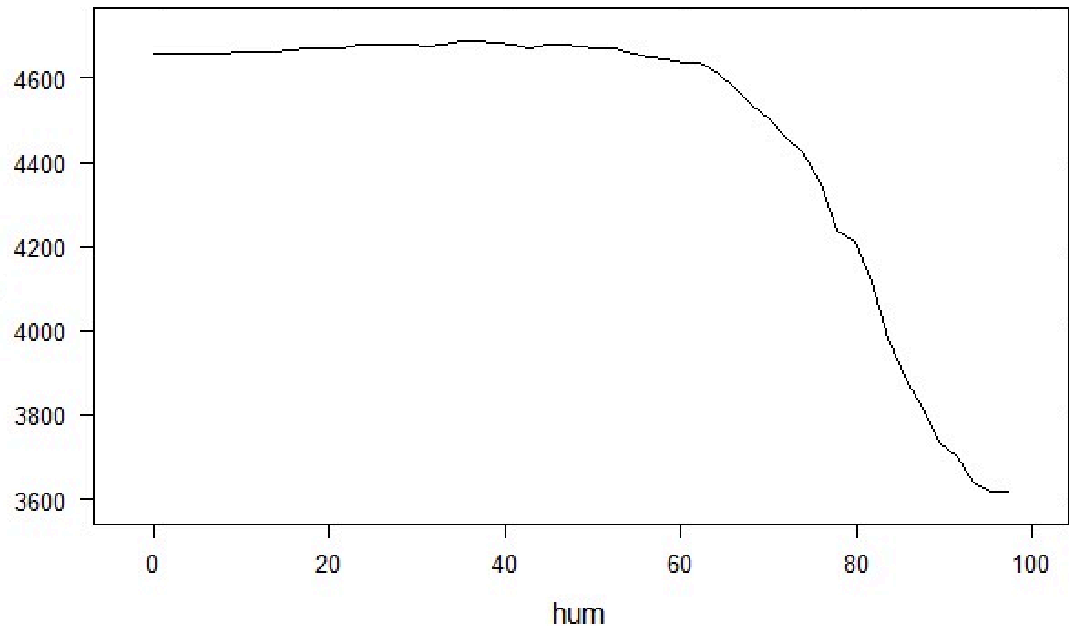


Figure 4: PDP for humidity

Figure 4 shows the change in average number of predicted bike rentals in respect to the humidity. For humidity levels from 0 to around 60 percent the number of predicted bikes is quite stable. From about 60 percent onward the predicted bike rentals fall drastically by around 1000 bikes until around 100 percent. This shows that up to a certain humidity level, humidity does not seem to play a role, but from high humidity onward it becomes more important.

Bidimensional Partial Dependence Plot

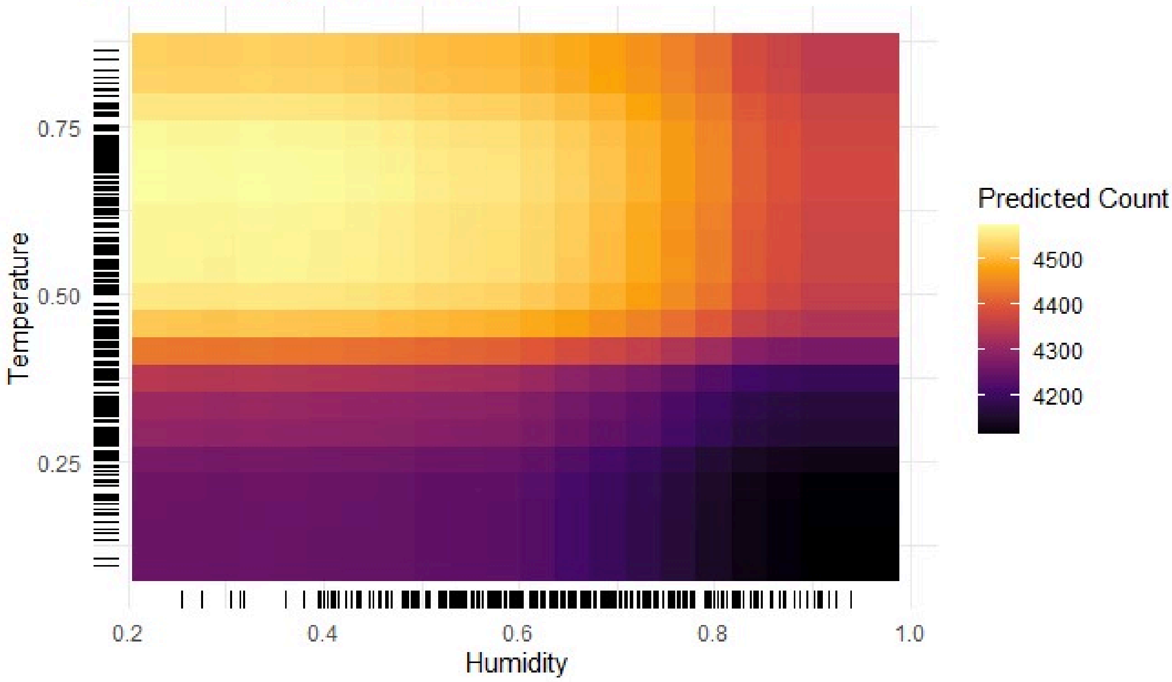


Figure 5: Bidimensional PDP for humidity and temperature

The bidimensional PDP shows the influence of humidity and temperature together on the bike rental predictions. Here the color represents the impact on the predictions. Darker colors indicate lower predictions of around 4000 bike rentals while lighter colors represent predictions of about 4600 bike rentals. This allows us to investigate the interaction of two variables (here: humidity and temperature).

This PDP reveals that the combination of higher temperatures and medium humidity are correlated with higher predictions of bike rentals. Higher humidity is generally connected to slightly lower rental predictions while low humidity does not seem to have a significant effect. The lowest rental predictions are seen in the low temperatures. In the low temperatures higher humidity also has a negative influence on predicted bike rentals while low humidity does not appear to be significant.

However, one has to keep in mind that only few data points are available to support the interpretation for low humidity values. The same goes for really high humidity, really low temperatures and really high temperatures. Therefore, the interpretations in those areas have to be taken with caution.

However, it is interesting to see that the influences of temperature and humidity in this plot pretty much reflect just the information we already saw in the individual PDPs of humidity and temperature. This could mean that there are no impactful interactions between temperature and humidity that affect the predictions in an unexpected way. In conclusion, one does not see much new information in contrast to the information from the two individual PDPs for these specific variables.

PDP to explain the price of a house

For this exercise, we have applied the previous concepts to predict the price of a house from our database using a Random Forest approximation.

We have used as features the number of bedrooms, bathrooms, square feet of living space, the square feet of lot, the floors and the year it was built.

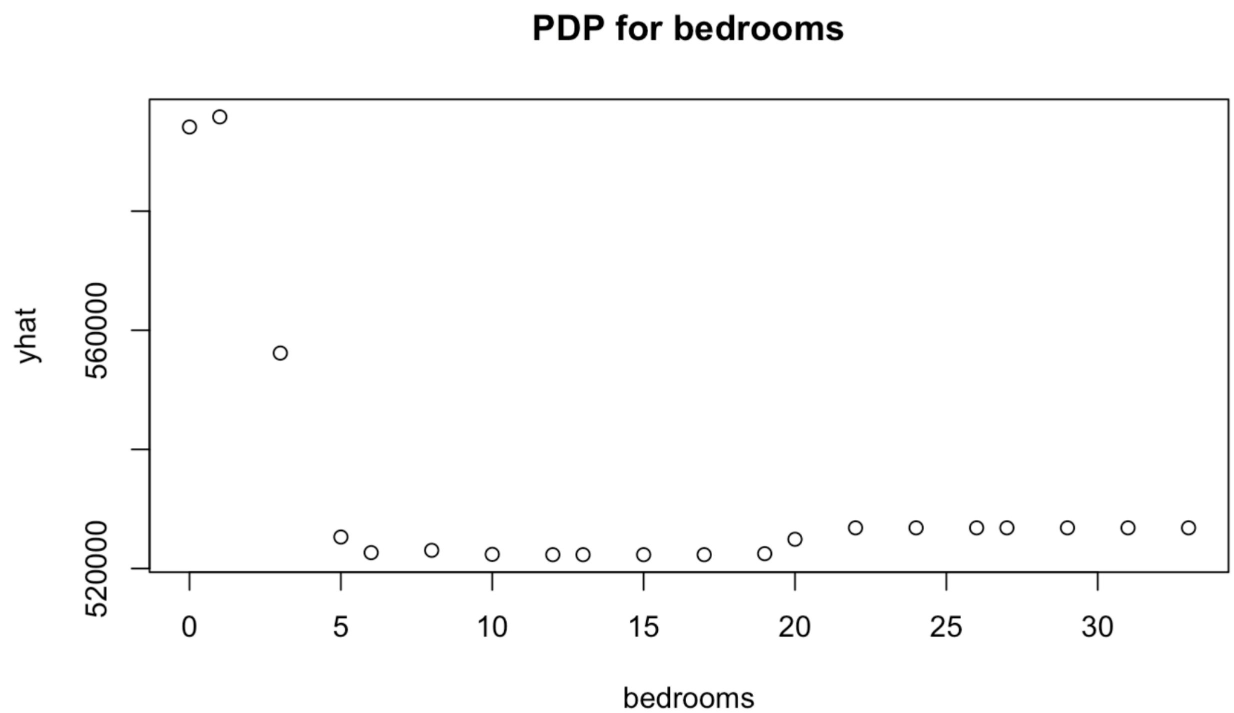
We are analyzing the influence of bedrooms, bathrooms, sqft_living and floors on the predicted price.

The Partial Dependency Plots (PDPs) obtained offer a visual representation of how different features impact the predicted price of a house.

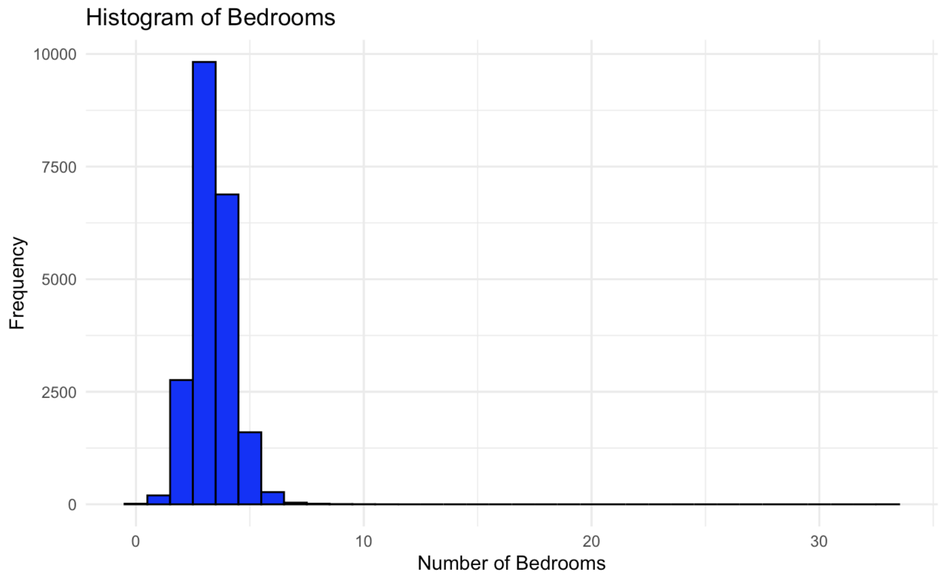
- **Bedrooms:** This plot shows a relatively flat trend across the number of bedrooms, with several outliers showing houses with an extreme number of bedrooms (around 30).

  Looking at the plot obtained, we notice that houses with more than 5 bedrooms don´t seem to follow a coherent distribution, obtaining, with a high probability, non representative values.

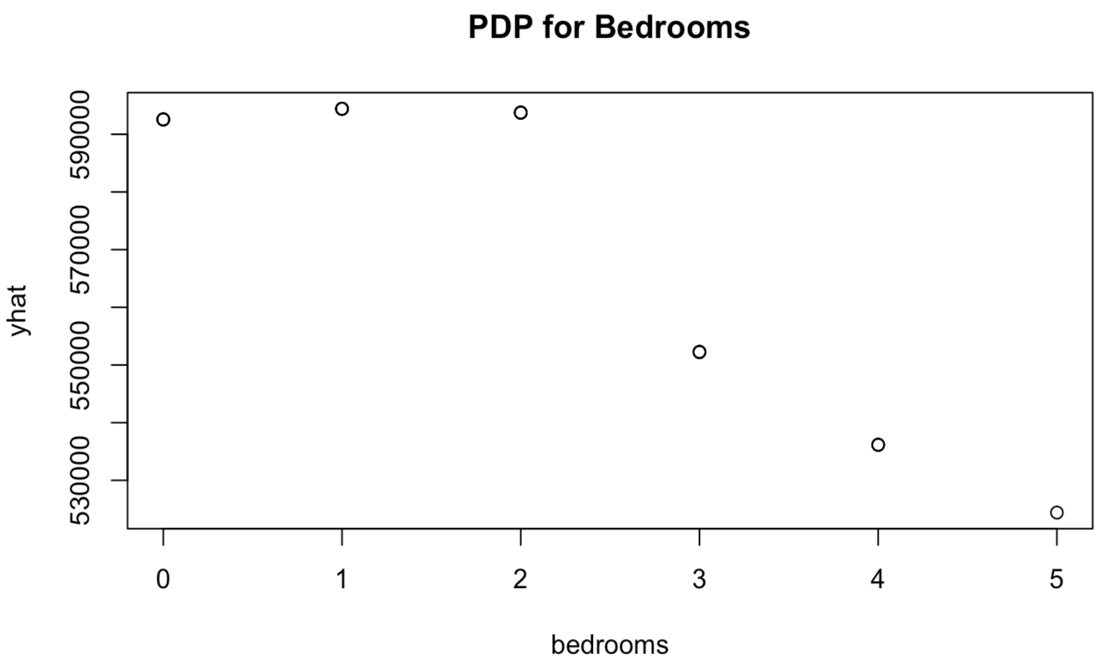
  For practical applications, the influence of bedrooms on house prices appears stable across typical values (1-10 bedrooms).



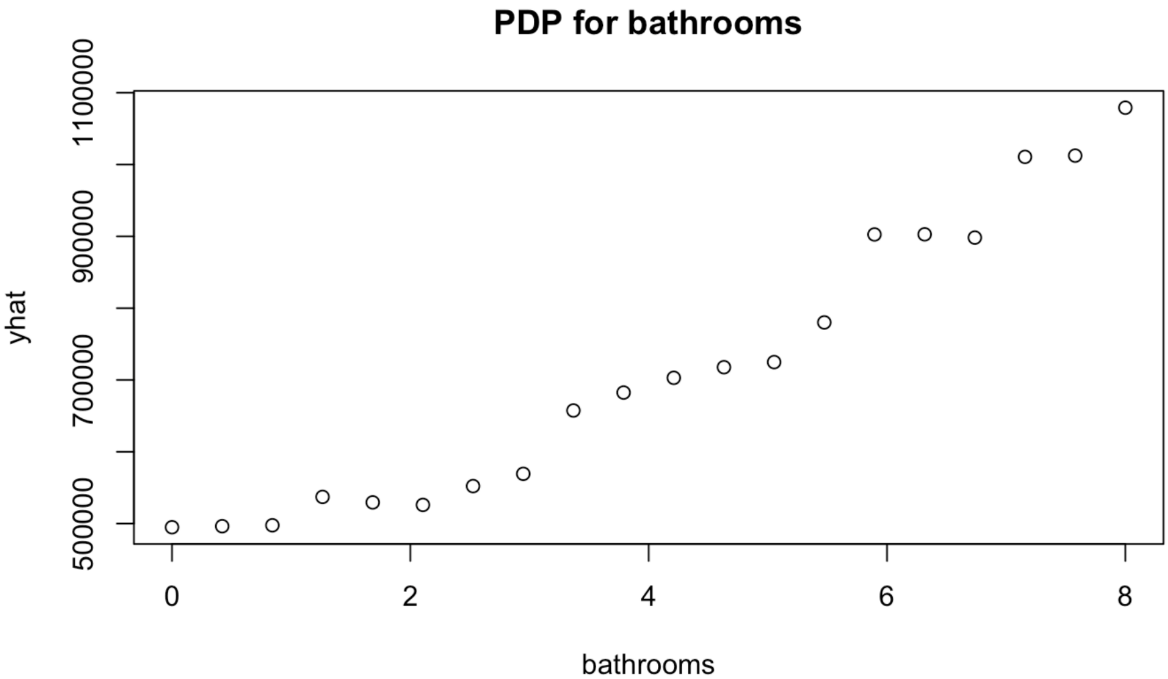
Filtering the data to get only houses with a maximum of 5 bedrooms, we can see this distribution:



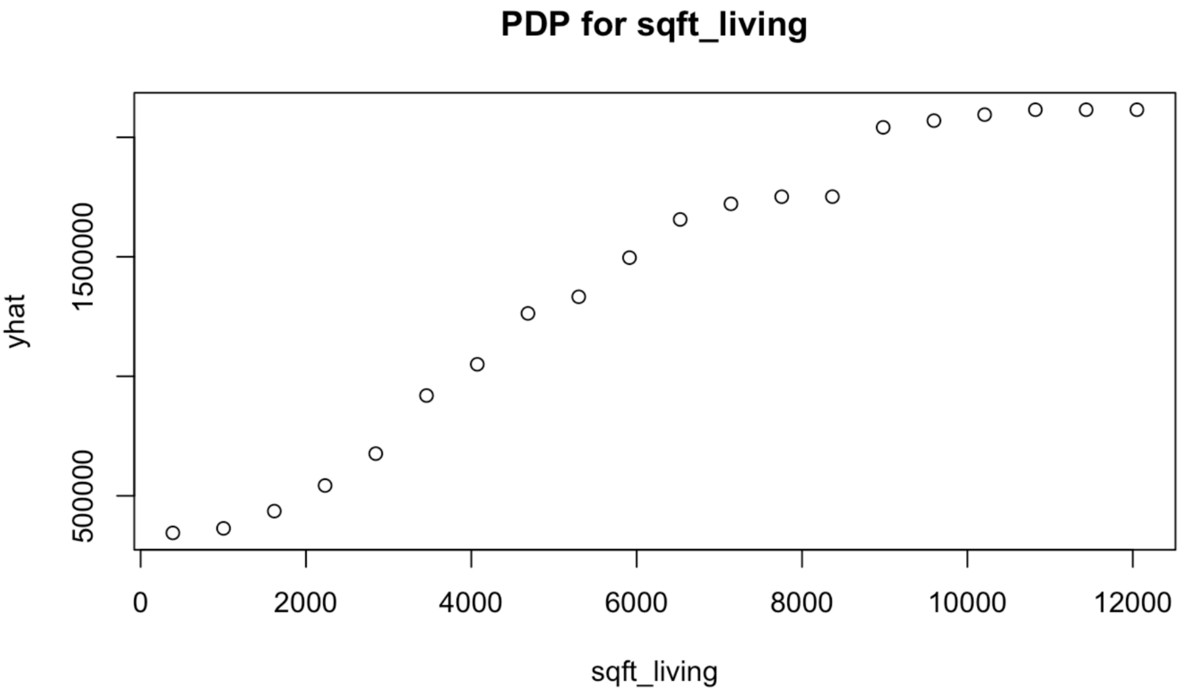
So we still don't find coherent the distribution where from 2 bedrooms in advance, the price of the house decreases significatively, agreeing on the point that those values are not representative.



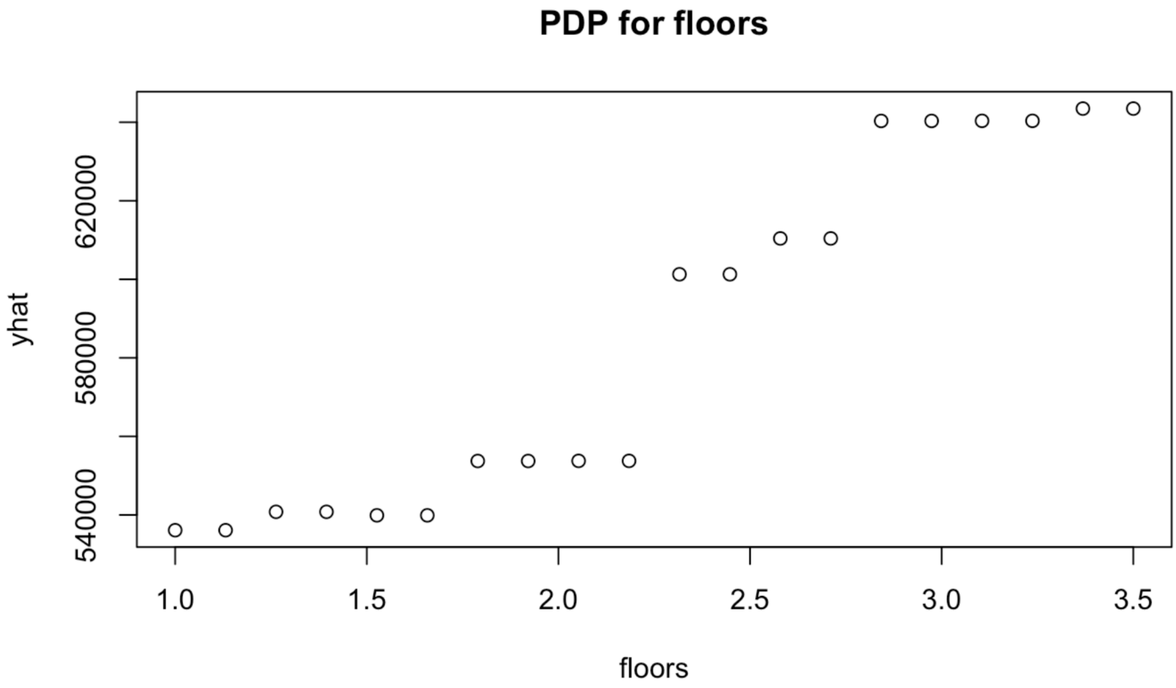
- **Bathrooms:** The data demonstrates that as the number of bathrooms in the houses increases so does the price, suggesting a direct relationship between the two. The most significant increase happens up to approximately 2 bathrooms, flattening out before rising again past 6 bathrooms.
  This indicates that having more bathrooms increases the value of a house, particularly in larger homes.



- **Sqft living:** There's a clear positive relationship between the living area size (sqft_living) and the house price. As the square footage of the living area increases, the price increases almost linearly. This highlights that larger living spaces are highly valued, which is consistent with real estate market expectations where more living space typically commands a higher price.



- **Floors:** The relationship on this point is less clear, with price showing slight increases as the number of floors increases, but not as dramatically as with sqft_living. The graph indicates a moderate influence of having more floors on increasing house prices, which may be valued for the views or separation



of living spaces they offer.

Conclusion

In this report, we have employed Partial Dependence Plots (PDP) as a model-agnostic explainability method to analyze the predictions made by Random Forest models for two distinct applications: bike rental predictions and house price predictions.

For the bike rental predictions, PDPs were used to assess the impact of various features including temperature, wind speed, days since 2011, and humidity. The analyses revealed that:

- **Temperature**: Bike rentals increase with temperature up to a peak at around 22 degrees Celsius, beyond which the predicted rentals decline slightly.

- **Wind Speed**: Higher wind speeds correlate with fewer predicted bike rentals, with a notable decrease starting at 5 kph.

- **Days Since 2011**: Bike rentals show a general increasing trend over time, with periodic fluctuations.

- **Humidity**: Predictions are stable up to 60% humidity, after which higher humidity levels significantly reduce bike rental numbers.

The bidimensional PDP for humidity and temperature corroborated these findings, indicating no significant interaction effects beyond those already observed in the individual PDPs.

For the house price predictions, PDPs were utilized to evaluate the influence of features such as the number of bedrooms, bathrooms, square footage of living space, and number of floors. The findings include:

- **Bedrooms**: A stable influence on house prices up to 5 bedrooms, with data for houses having more than 5 bedrooms appearing less reliable.

- **Bathrooms**: A direct positive relationship with house prices, especially significant up to 2 bathrooms and again beyond 6 bathrooms.

- **Living Space (Sqft Living)**: A clear positive linear relationship, with larger living areas commanding higher prices.

- **Floors**: A moderate influence, with more floors slightly increasing house prices, likely due to the benefits of views and space separation.

In conclusion, PDPs have proven to be an effective tool for interpreting model predictions, providing valuable insights into how different features influence the outcomes. They allow for a better understanding of feature importance and interactions, thereby enhancing the transparency and trustworthiness of complex machine learning models. For future work, it would be beneficial to explore additional model-agnostic methods and apply them to other types of models to further validate and enrich these findings.