XAI 3

José Aguilar Camps & Joel Moncho Mas

MAY 2025

Contents

1 Introduction **2**

2 One dimensional Partial Dependence Plot **2**

2.1 Data Loading and Preprocessing . . . 3

2.2 Model Training and PDP Generation . . . 3

2.3 Interpretation of the PDP: Days since 2011 vs. Predicted Bike Rentals . . . 4

2.4 Interpretation of the PDP: Temperature vs. Predicted Bike Rentals . . . 5

2.5 Interpretation of the PDP: Humidity vs. Predicted Bike Rentals . . . 6

2.6 Interpretation of the PDP: Wind Speed vs. Predicted Bike Rentals . . . 7

3 Bidimensional Partial Dependency Plot **7**

3.1 Explanation of the Code . . . 8

3.2 Interpretation of the 2D PDP: Temperature vs. Humidity for Predicted Bike Rentals . . . 9

4 PDP for House Price Prediction **10**

4.1 PDP for num_bedrooms . . . 14

4.2 PDP for num_bathrooms . . . 14

4.3 PDP for sqft_living . . . 15

4.4 Final Conclusions . . . 15

Note: The RMD code for this document is available in our Github repository.

1 Introduction

This report presents an analysis based on real-world datasets, focusing on the application of **Partial Dependency Plots (PDPs)** to explore the relationships between features and predicted outcomes. Specifically, we will work with the **Capital Bikeshare dataset**, which provides detailed data on bike rentals in Washington D.C. over the years 2011 and 2012. Additionally, we will apply this analysis to predict bike rental volume and house prices.

The main objective of this study is to construct interpretable models capable of predicting the following outcomes: predicting the number of bikes rented based on various weather and time-based features such as temperature, humidity, and other environmental variables, and analyzing how different housing characteristics, including the number of bedrooms, bathrooms, and the size of the house, influence the predicted price of a house.

To achieve this, we apply the **Partial Dependency Plot (PDP)** technique to visualize the marginal effect of features on the target variables (bike rentals and house price), making it easier to understand the relationships between features and predictions.

We will be working with two primary datasets in this analysis. The first dataset is the **Capital Bikeshare dataset**, which includes hourly bike rental data from Washington D.C. for the years 2011 and 2012. The target variable in this dataset is the number of bikes rented (denoted as `cnt`), and the features include weather-related data such as temperature, humidity, and wind speed, along with time-based features like weekdays, holidays, and seasons.

The second dataset is the **House Price dataset**, which includes features like the number of **bedrooms**, **bathrooms**, **square footage** of living and lot area, **number of floors**, and **year built**, used to predict the **price** of houses. In this case, the goal is to understand how these features contribute to estimating house prices using PDPs.

The focus of the analysis is not only on building models that provide accurate predictions but also on **interpretability**. By visualizing the relationships between key features and the predicted outcomes, we aim to identify how changes in certain features (such as temperature or the number of bedrooms) affect the target variable (bike rental count or house price). We will explore the **marginal effects** of each feature using PDPs and provide **insights** into how the model makes decisions, thus making it more transparent and interpretable, especially for decision-making contexts.

Through the use of **Partial Dependency Plots**, we aim to visualize **non-linear relationships** and feature interactions, ensuring that the models are not only accurate but also understandable. Ultimately, the goal is to create transparent models that can provide clear explanations to non-technical users about how predictions are made.

In the following sections, we will generate PDPs for bike rentals and house prices, analyze the results, and discuss the insights gained from these visualizations.

2 One dimensional Partial Dependence Plot

In this section, we perform a series of steps to process the data and generate **Partial Dependency Plots (PDPs)** for different variables that influence the number of **bike rentals**. The goal is to understand how each feature, such as **days since 2011**, **temperature**, **humidity**, and **wind speed**, affects the predicted number of bike rentals.

2.1 Data Loading and Preprocessing

First, we load the **Capital Bikeshare dataset** using `read_csv("day.csv")`. This dataset contains hourly records of bike rentals, with variables like temperature, humidity, and other weather-related features. We apply several preprocessing steps to transform the data and make it suitable for model training and interpretation.

1. **Seasonal Variables**: We create binary variables for the seasons (`spring`, `summer`, `fall`) based on the `season` variable. This transformation allows us to better capture the seasonal effects on bike rentals. For instance, spring (`season == 1`) is marked as `spring = 1`, and other seasons are marked as `0`.

2. **Weather Variables**: We also create binary variables to indicate whether the weather was misty (`MISTY`) or rainy (`RAIN`). These new variables are derived from the `weathersit` feature, where we treat `weathersit == 2` as misty and `weathersit == 3` or `weathersit == 4` as rainy. These weather conditions likely impact bike rental behavior, so it's crucial to isolate these effects.

3. **Feature Scaling**:

   - **Temperature (`temp`)**: The temperature is scaled to actual Celsius values by multiplying by `41`.
   - **Humidity (`hum`)**: Humidity is scaled to a percentage by multiplying by `100`.
   - **Wind Speed (`windspeed`)**: The wind speed is scaled to kilometers per hour (km/h) by multiplying by `67`.

   These transformations help to make the features more interpretable in real-world terms.

4. **Date Transformation**: The `dteday` variable, which represents the date, is converted into a **Date** object using `as.Date()`. Additionally, we calculate **days since 2011** using `difftime(dteday, as.Date("2011-01-01"), units = "days")`, which allows us to quantify the temporal aspect of bike rentals and study the trend over time.

After preprocessing, we select the relevant features for the model, including the target variable (`cnt` for bike rentals), weather variables, and scaled features. This dataset is now ready for model training.

2.2 Model Training and PDP Generation

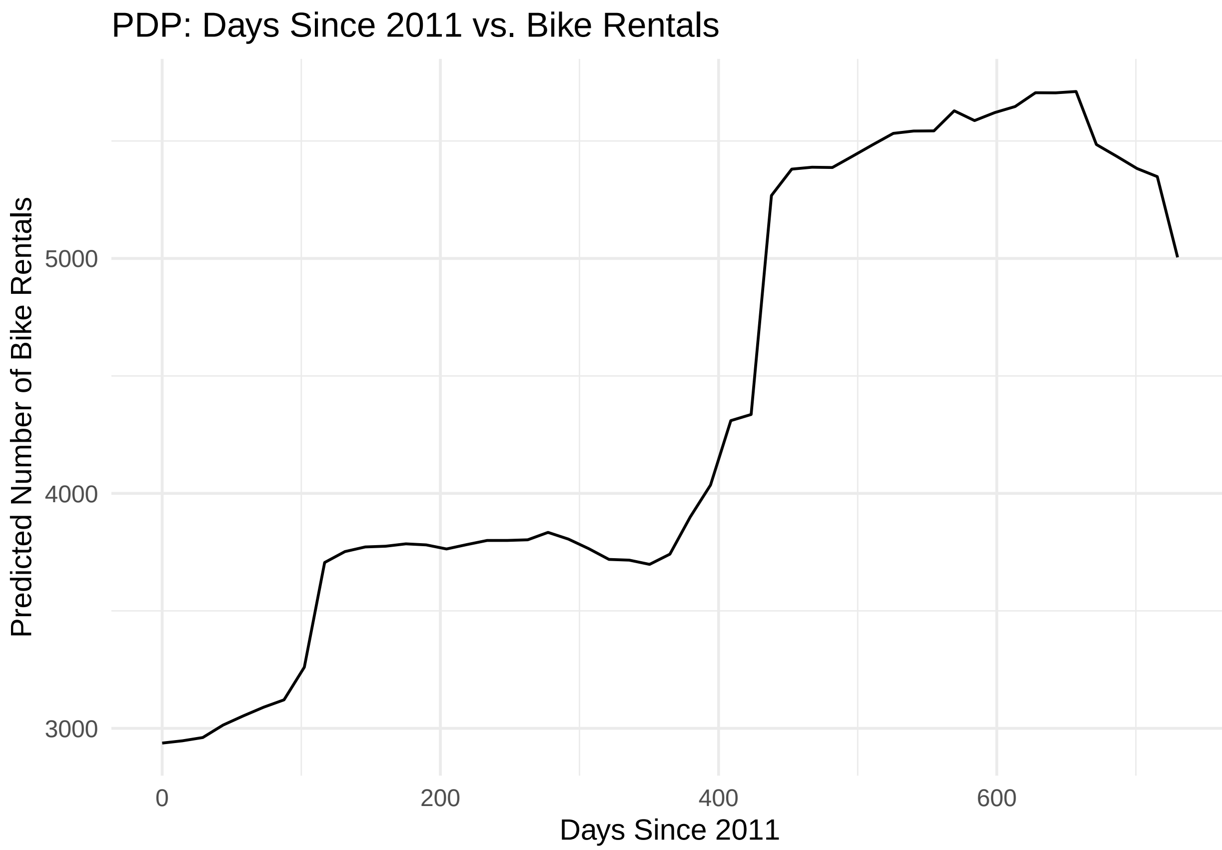
We train a **Random Forest** model (`rf_model`) to predict the number of bike rentals (`cnt`) using the preprocessed features. The model is trained on variables such as **temperature**, **humidity**, **wind speed**, and **days since 2011**.

Once the model is trained, we apply **Partial Dependency Plots (PDPs)** to visualize the influence of each feature on the predicted number of bike rentals. For each feature, we generate a PDP using the `partial()` function, which computes the predicted outcome while varying the selected feature across its range, holding all other features constant.

We generate and plot the PDPs for each of the features (**days since 2011**, **temperature**, **humidity**, and **wind speed**) using **`ggplot2`**. Each plot shows how the predicted number of bike rentals changes as a single feature varies, helping us understand the model's behavior and the relationships it has learned between the features and the target variable.

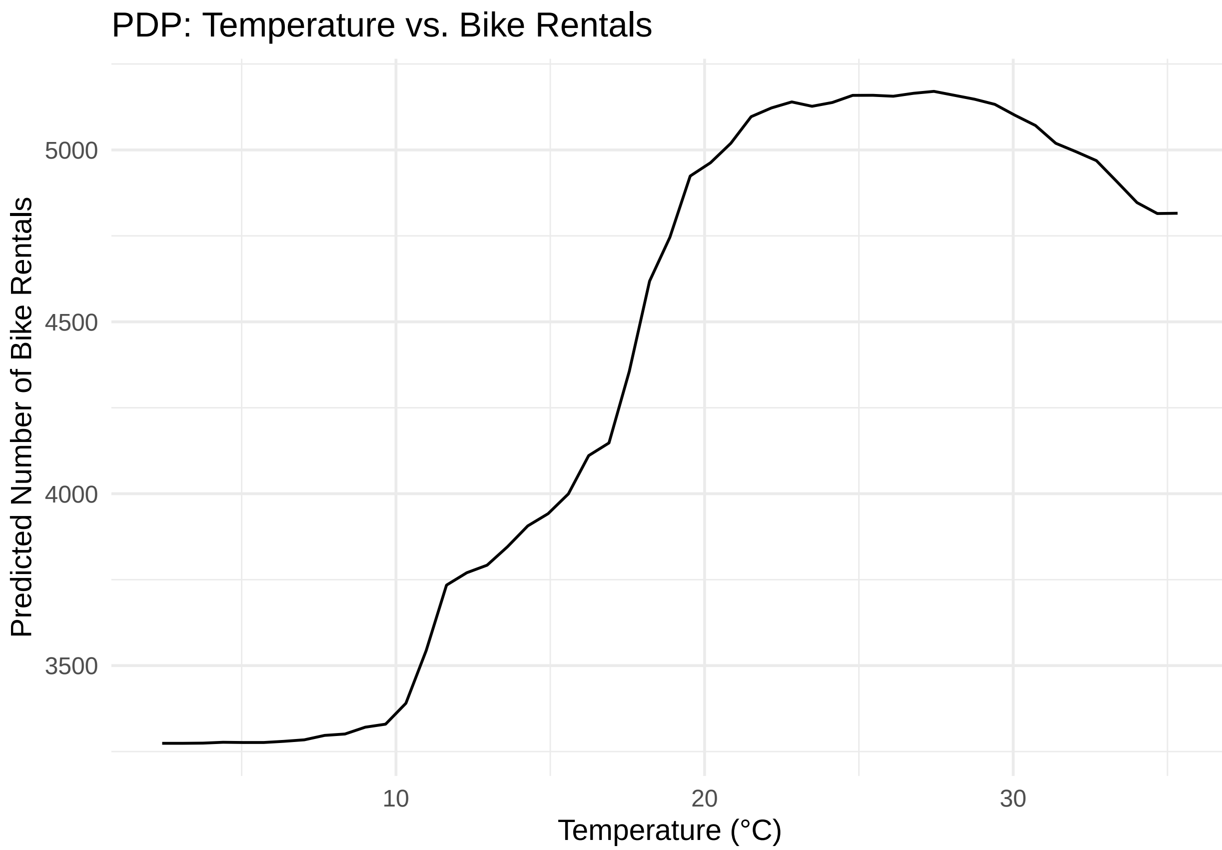
In the following visualizations, we will interpret how each feature, individually, influences the predicted number of bike rentals, and gain insights into the most important variables for understanding rental behavior.

2.3 Interpretation of the PDP: Days since 2011 vs. Predicted Bike Rentals



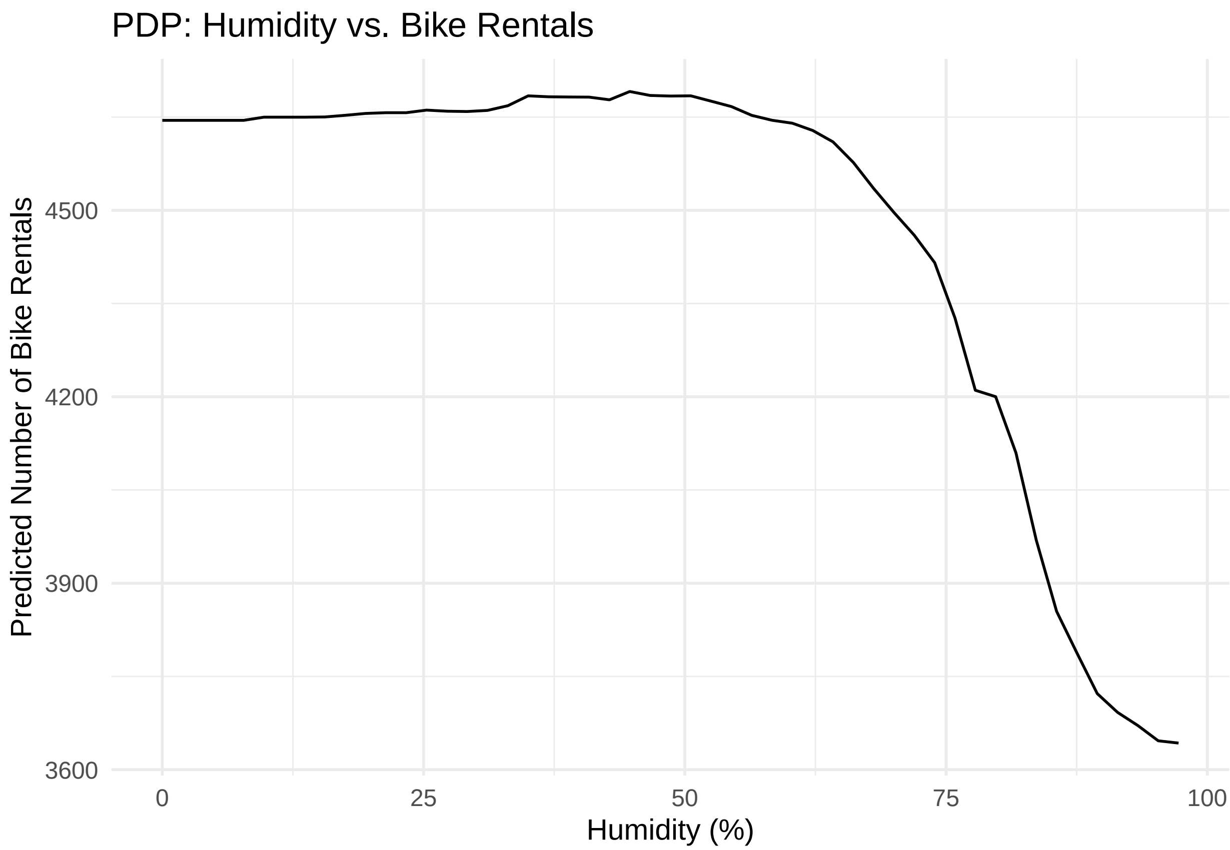
The "Days Since 2011 vs. Predicted Bike Rentals" plot shows how the number of bike rentals changes over time, starting from the beginning of 2011. Initially, when the number of days is close to zero, the predicted rentals are low—likely due to the colder weather in the early months of the year. As time progresses, particularly after day 200, there is a noticeable increase in predicted rentals, which likely reflects the onset of warmer seasons and more favorable conditions for cycling. The number of rentals then reaches a plateau, suggesting that demand stabilizes once the weather becomes consistently pleasant. Toward the end of the period (around day 600), a slight decline in predicted rentals appears, possibly due to the end of peak season, where colder temperatures or less favorable conditions begin to reduce bike usage.

2.4 Interpretation of the PDP: Temperature vs. Predicted Bike Rentals



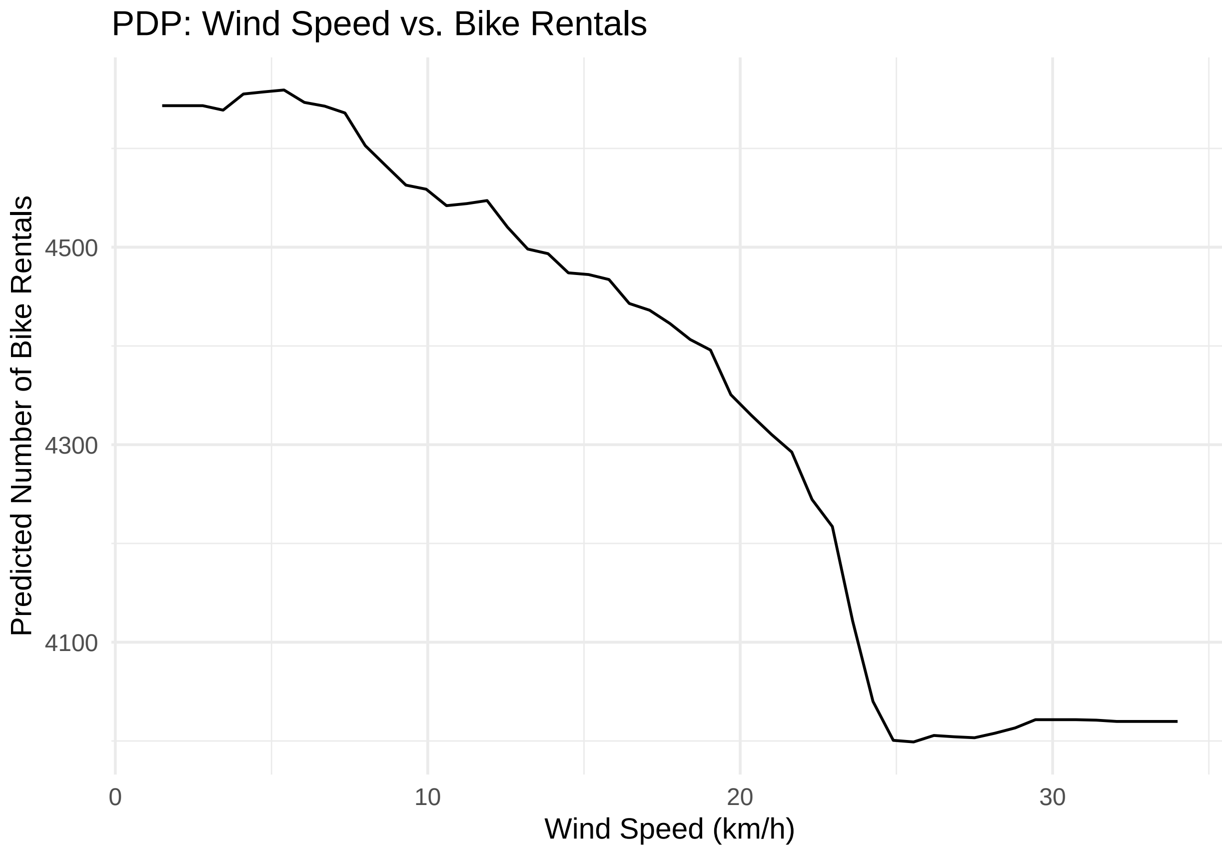
The **"Temperature vs. Predicted Bike Rentals"** plot shows how the predicted number of bike rentals varies with temperature. As the temperature begins to rise from lower values, there is a sharp increase in predicted bike rentals, indicating that people are more inclined to rent bikes when the weather becomes more comfortable. This positive correlation is expected, as warmer temperatures generally promote outdoor activities like cycling. Around 20°C, the curve begins to flatten, suggesting that beyond this point, further increases in temperature have a diminishing effect on rental demand. Interestingly, as the temperature exceeds 25–30°C, the plot shows a decline in predicted rentals. This suggests that excessively hot conditions may discourage people from biking due to discomfort or health risks. Overall, the plot highlights an optimal temperature window for bike rentals, with peak usage observed roughly between 20°C and 25°C.

2.5 Interpretation of the PDP: Humidity vs. Predicted Bike Rentals



The **"Humidity vs. Predicted Bike Rentals"** plot shows how the predicted number of bike rentals varies with humidity levels. At low to moderate humidity (up to around 60%), the number of predicted rentals remains relatively stable, indicating that these conditions do not significantly affect people's likelihood to rent bikes. However, beyond approximately 65–70% humidity, the predicted number of rentals begins to decline sharply. This suggests that high humidity levels make cycling less appealing, likely due to discomfort caused by muggy or oppressive conditions. At very high humidity levels (above 80%), predicted bike rentals drop substantially, highlighting how excessive humidity can deter outdoor activities. Overall, the plot indicates that low to moderate humidity is optimal for bike rentals, while high humidity acts as a clear deterrent.

2.6 Interpretation of the PDP: Wind Speed vs. Predicted Bike Rentals



The **"Wind Speed vs. Predicted Bike Rentals"** plot illustrates how the predicted number of bike rentals changes as wind speed increases. Initially, the number of predicted rentals remains relatively stable as wind speed rises from 0 km/h to around 10 km/h, suggesting that light winds have little impact on people's willingness to rent bikes. However, once wind speeds exceed 10 km/h, the predicted rentals begin to gradually decline, indicating that stronger winds may start to affect comfort and perceived safety. This downward trend becomes more pronounced beyond 20 km/h, with a sharp drop in predicted bike rentals occurring after approximately 30 km/h. This suggests that high wind speeds significantly discourage bike usage, likely due to physical difficulty or safety concerns. Overall, the plot highlights that mild to moderate wind conditions are generally acceptable for bike rentals, while strong winds serve as a clear deterrent.

---

3 Bidimensional Partial Dependency Plot

In this exercise, we will generate a **2D Partial Dependency Plot (PDP)** to analyze how **temperature** and **humidity** jointly affect the predicted number of bike rentals. The goal is to understand the interaction between these two features and their combined influence on bike rental behavior. To manage the dataset's size, we will first extract a random sample of data from the database before generating the PDP. Additionally, we will visualize the **density distribution** of both input features—**temperature** and **humidity**—using the 2D plot, as shown in the class slides. The **geom_tile()** function will be used to create the 2D plot, ensuring that the width and height of the tiles are adjusted to avoid any gaps in the plot.

This exercise will help us interpret how the relationship between temperature and humidity influences bike rental predictions, providing a deeper understanding of the model's behavior.

3.1 Explanation of the Code

In this code, we begin by **loading and preprocessing the data**. To ensure that the dataset is manageable and computationally efficient, we randomly sample 100 rows from the **Capital Bikeshare dataset** using `sample_n()`. We then select the relevant columns, which include the **bike rentals count (`cnt`)**, **temperature (`temp`)**, and **humidity (`hum`)**.

Next, we apply **scaling** to the **temperature** and **humidity** variables to convert them into more interpretable units. The temperature is scaled to actual **Celsius** values by multiplying by `41` (since the original temperature is in a 0-1 range), and humidity is scaled to **percent** by multiplying by `100`. This scaling makes the variables more meaningful for interpretation in real-world terms.

3.1.1 Why Random Forest and PDP?

For the model, we choose a **Random Forest** approach (`randomForest()`), as it is a robust model that can handle non-linear relationships between features and the target variable. This is important because the relationship between **temperature**, **humidity**, and **bike rentals** is likely to be complex, and Random Forest can capture these interactions effectively. We then use the **Partial Dependency Plot (PDP)** technique to visualize how the predicted bike rentals change as **temperature** and **humidity** vary.

We apply the **`partial()`** function to compute the PDP for the two features (`real_temp` and `real_hum`). The `plot = FALSE` argument ensures that we generate the PDP data without displaying it immediately.

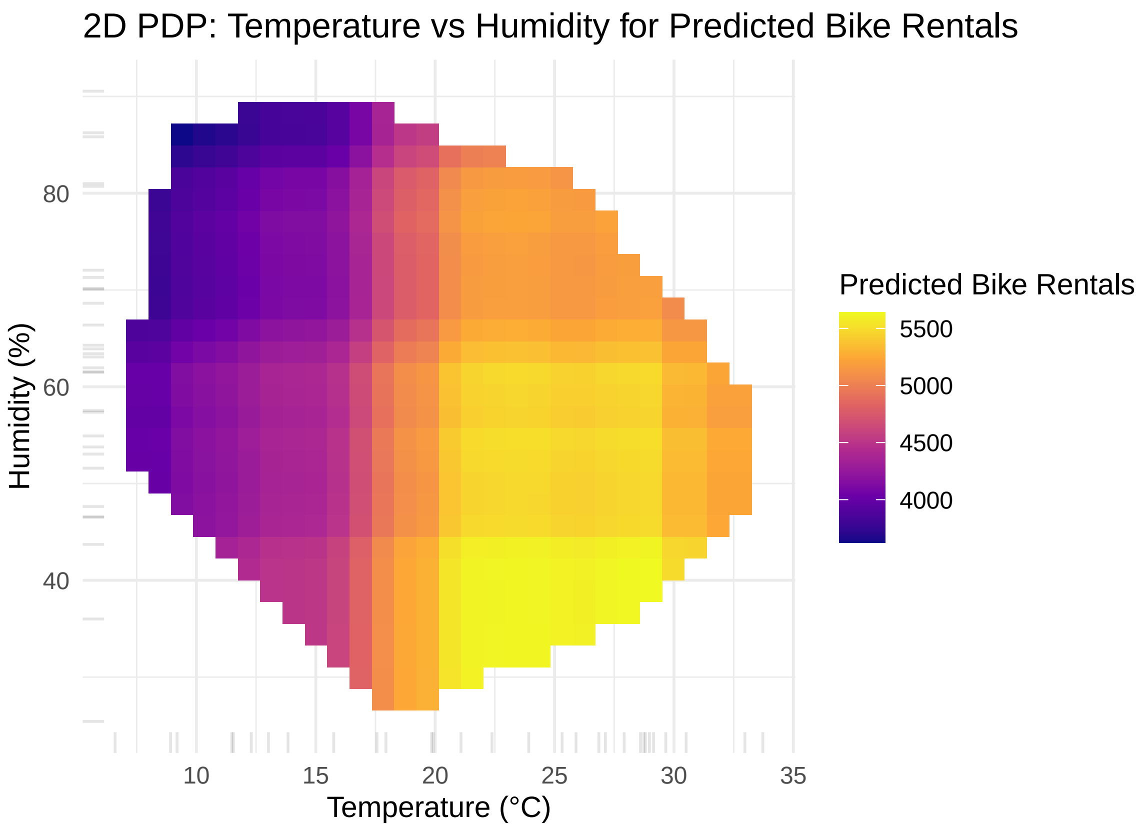
3.1.2 Calculating Tile Width and Height

To create the **2D plot** using **`geom_tile()`**, we first determine the size of the tiles. The size is determined by the unique values of **temperature** and **humidity** in the dataset. We calculate the **tile width** and **tile height** based on the difference between the unique values of each feature. If there is only one unique value, we provide a fallback value to ensure the plot is still generated without any gaps. This step ensures the plot is smooth and accurately represents the data.

3.1.3 Generating the 2D PDP

Finally, we use `ggplot2` to generate the **2D PDP**. The `geom_tile()` function is used to create the heatmap, where the **x-axis** represents **temperature**, the **y-axis** represents **humidity**, and the **fill** indicates the predicted number of bike rentals (`yhat`). The **`scale_fill_viridis_c()`** function is used to color the plot, providing a clear visual gradient to indicate predicted bike rental counts. Additionally, we use **`geom_rug()`** to add **density distribution lines** for both temperature and humidity on the plot, helping to visualize the concentration of data points.

The final plot is displayed with **minimalist themes** and labeled appropriately for clarity.



3.2 Interpretation of the 2D PDP: Temperature vs. Humidity for Predicted Bike Rentals

The **2D Partial Dependence Plot (PDP)** visualizes how the predicted number of bike rentals is influenced by the interaction between **temperature** and **humidity**. This plot helps us understand how changes in these two features, in combination, affect bike rental predictions, providing insights into their joint influence on the target variable.

3.2.1 Key Observations

- **Impact of Temperature**: As expected, the plot shows a clear **increase in bike rentals as the temperature rises**, with rentals becoming more frequent as temperatures exceed approximately 15°C. This trend suggests that higher temperatures create more favorable conditions for cycling, leading to more people renting bikes.
- **Effect of Humidity**:
  - At **lower humidity levels** (below 40-50%), there is a steady increase in bike rentals as the temperature rises. This indicates that when the air is not too humid, higher temperatures are positively associated with more bike rentals.
  - However, as **humidity increases** beyond 60-70%, the predicted number of rentals starts to **decrease**, even with higher temperatures. This trend suggests that **high humidity** deters people from renting bikes, likely due to discomfort or reduced willingness to engage in outdoor activities in humid conditions.

- **Peak Rental Activity**: The **highest predicted number of bike rentals** is observed when **temperature** is moderate (between 15°C and 25°C) and **humidity** is low to moderate (below 60%). This region is represented by the brightest yellow in the plot, indicating the most favorable conditions for bike rentals.
- **Decline in Rentals at High Humidity**: As the **humidity level rises** above 70%, a noticeable **decline in rentals** occurs, even at higher temperatures. This shows that **extremely humid conditions** significantly discourage bike rentals, likely because the discomfort of high humidity outweighs the appeal of cycling in warmer weather.
- **Interaction Between Temperature and Humidity**: The plot clearly demonstrates the **interaction between temperature and humidity**. While higher temperatures generally lead to more rentals, this effect is moderated by the level of humidity. The **best conditions for bike rentals** occur when both **temperature** and **humidity** are moderate, suggesting that a balance of these two factors is key to maximizing rental activity.

3.2.2 Conclusion

This **2D PDP** provides valuable insights into the relationship between **temperature** and **humidity** in predicting bike rentals. It highlights that **moderate temperatures** paired with **low to moderate humidity** lead to the highest predicted rental activity, while **high humidity** levels—especially in combination with higher temperatures—serve as a deterrent to bike rentals. Understanding these interactions can help in predicting bike rental demand based on weather conditions, providing a useful tool for bike share services to optimize their operations.

---

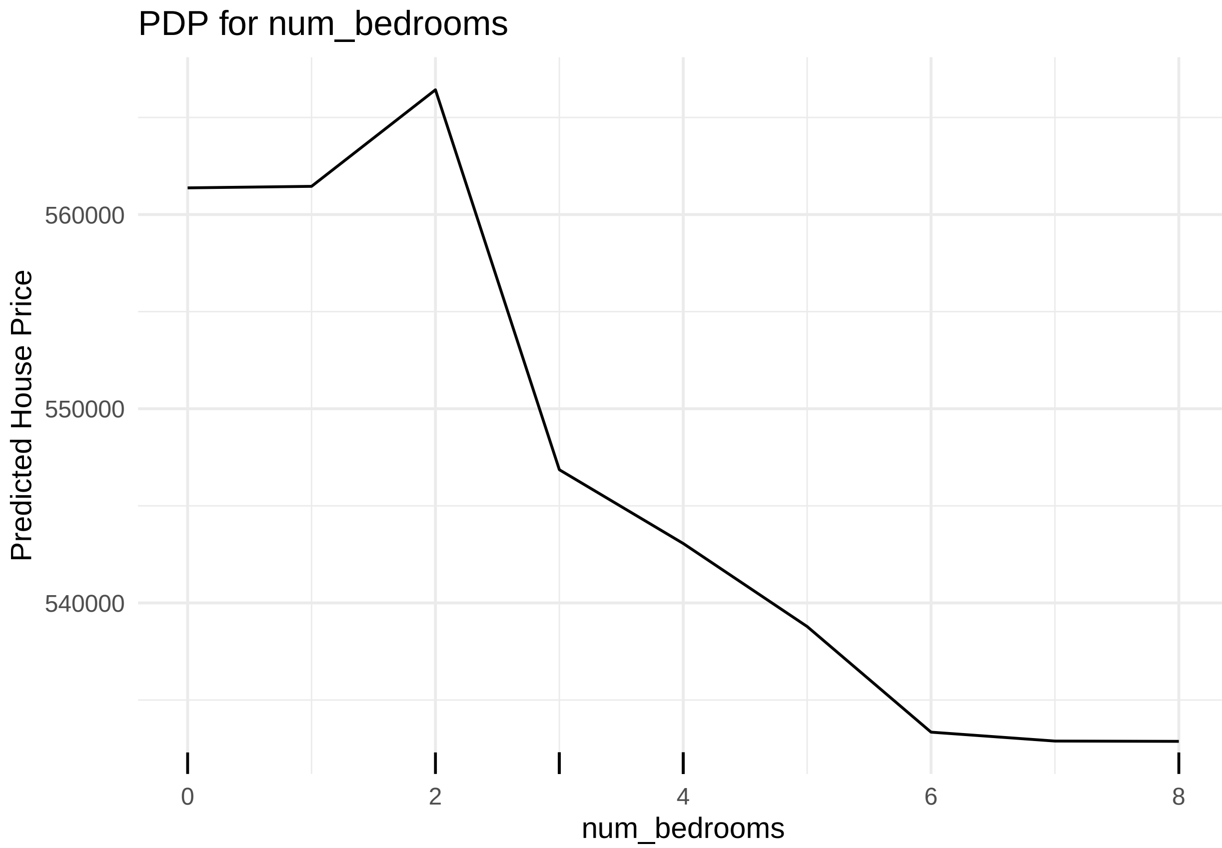
4 PDP for House Price Prediction

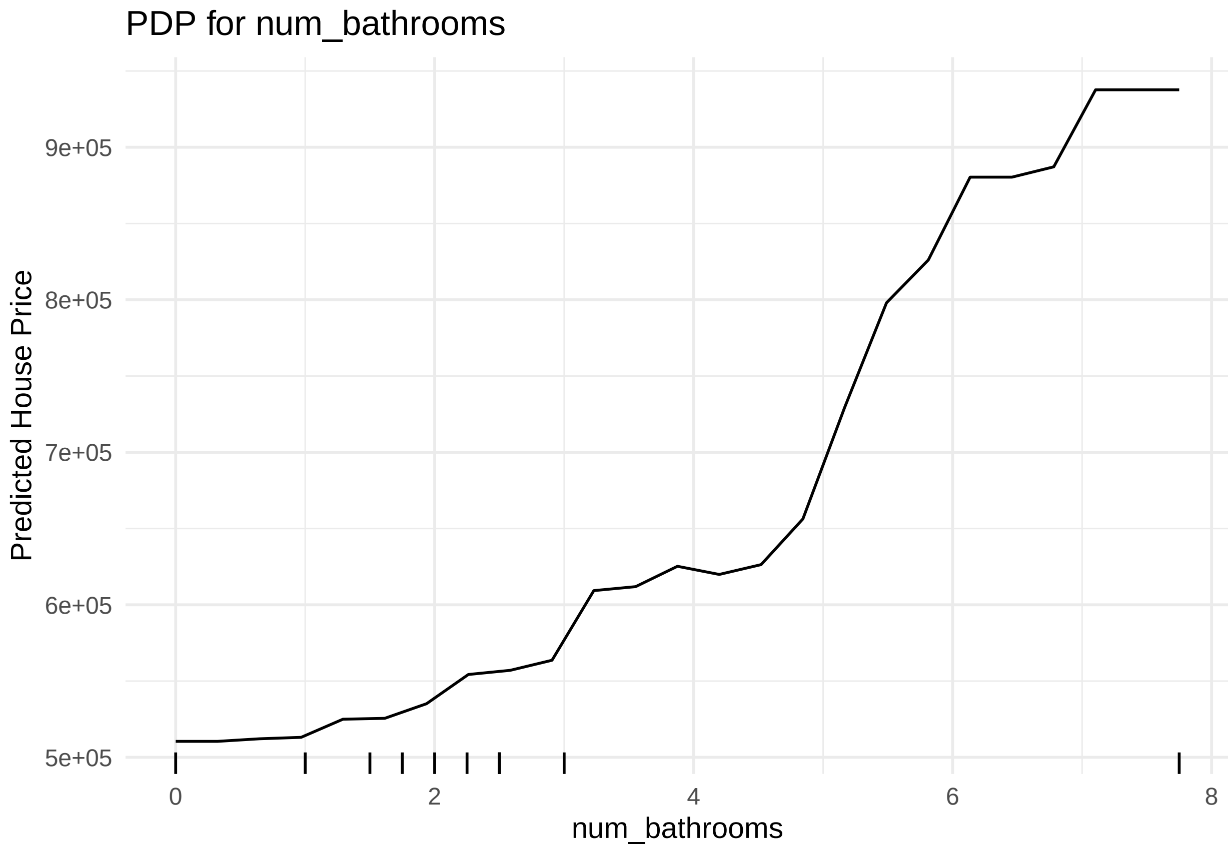
In this exercise, we will use **Partial Dependence Plots (PDPs)** to analyze how various features, such as **number of bedrooms**, **number of bathrooms**, **square footage of the living area**, and **number of floors**, influence the predicted price of a house from the **kc_house_data.csv** dataset. The aim is to apply the concept of **PDP** to visualize the relationships between these key features and the predicted price, helping us understand how the model interprets these factors.

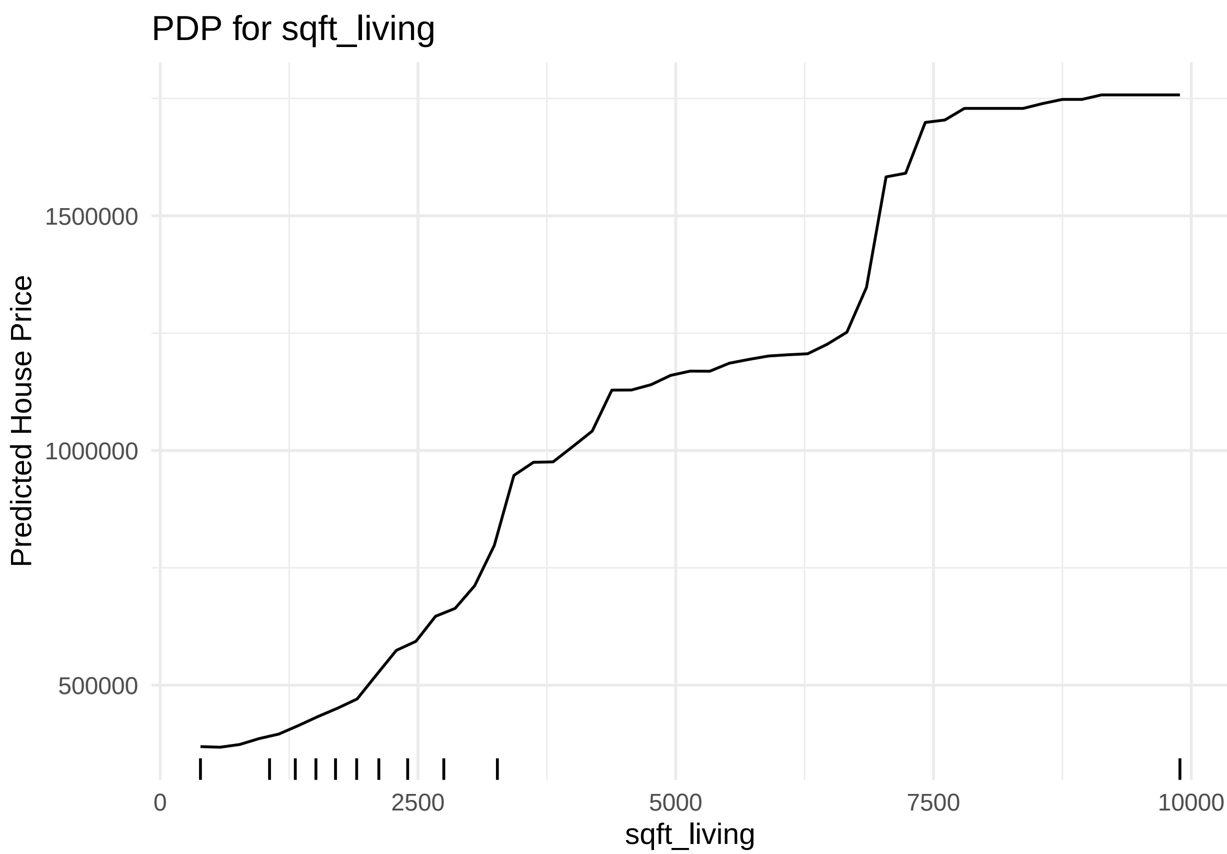
4.0.1 Data Loading and Preprocessing

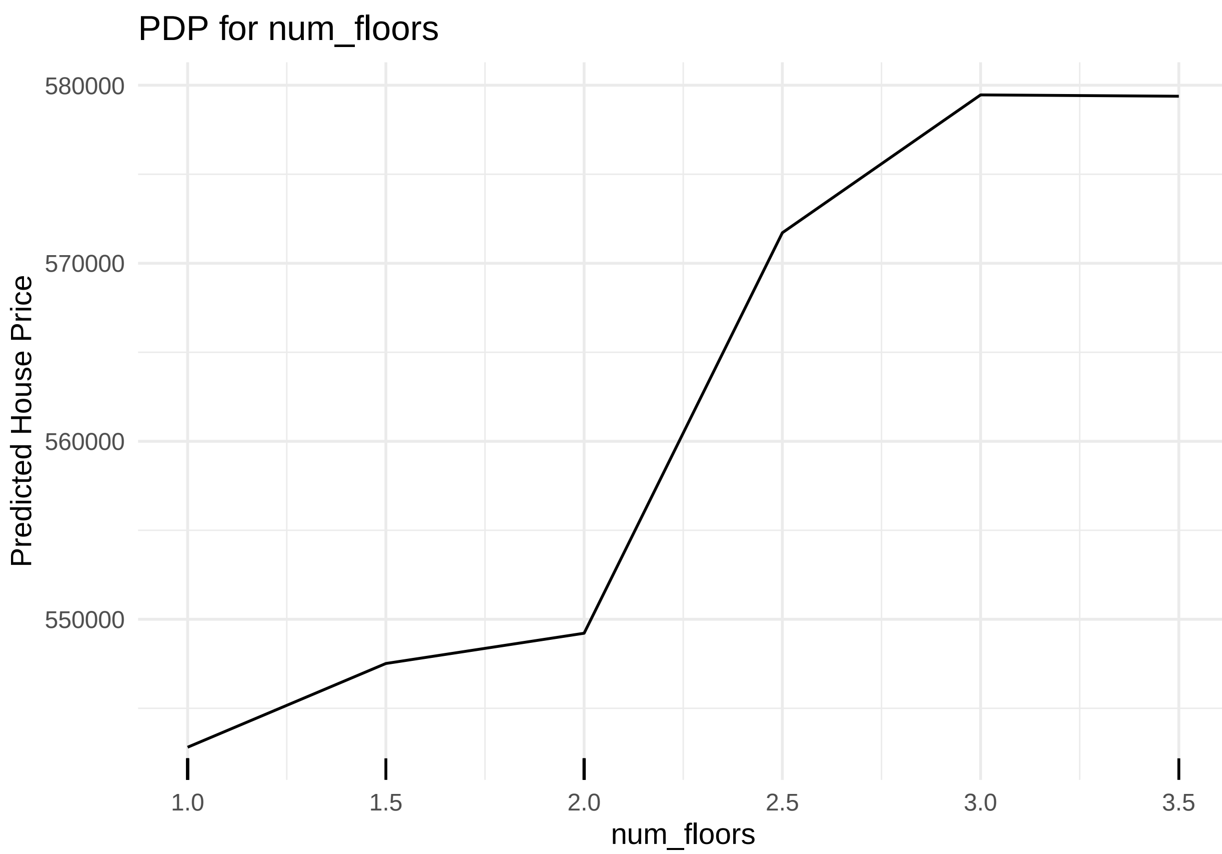
We begin by loading the **kc_house_data.csv** dataset, which contains various features related to houses such as the number of **bedrooms**, **bathrooms**, **square footage of the living area**, and others. We then select the relevant features for the model, which are the ones that have a direct influence on the house price, including **bedrooms**, **bathrooms**, **sqft_living**, **sqft_lot**, **floors**, and **yr_built**.

Next, the data is preprocessed: 1. **Feature Selection**: We select only the features relevant to our analysis. 2. **Missing Data**: We check for any missing values and remove any rows with missing data using `na.omit()`. 3. **Random Sampling**: Due to the size of the dataset, we randomly sample 5000 rows to keep the computation manageable. This ensures that we do not run into memory or computational issues while training the model.


PDP for num_bedrooms
Predicted House Price
560000
550000
540000
0
2
4
6
8
num_bedrooms


PDP for num_bathrooms
Predicted House Price
9e+05
8e+05
7e+05
6e+05
5e+05
0
2
4
6
8
num_bathrooms


PDP for sqft_living
Predicted House Price
1500000
1000000
500000
0
2500
5000
7500
10000
sqft_living



4.1 PDP for num_bedrooms

This plot shows how the predicted house price changes with the number of bedrooms. For homes with 1 to 2 bedrooms, the model predicts relatively high prices, around $560,000. However, as the number increases from 2 to 6, the predicted price drops significantly. This suggests that adding more bedrooms beyond a certain point may not add value and could even reduce it.

Interestingly, beyond 6 bedrooms, the predicted price appears to stabilize at a lower level. However, the support ticks indicate very few homes in this range, meaning predictions there are based on limited data and may be less reliable.

Overall, this PDP reveals a non-linear and somewhat counterintuitive relationship: more bedrooms do not necessarily increase the value of a home. Factors such as location, outdated layout, or poor space distribution may explain why larger homes are not always priced higher.

4.2 PDP for num_bathrooms

This PDP illustrates a strong and mostly monotonic positive relationship between the number of bathrooms and the predicted house price. From 0 to 3 bathrooms, the price rises gradually, indicating incremental value. Beyond 4 bathrooms, the price increases more steeply, suggesting that homes with 5 or more bathrooms are often high-end properties.

However, the data support drops significantly beyond 5 bathrooms, as shown by the sparse rug ticks. Therefore, predictions in that range should be interpreted with caution.

In summary, num_bathrooms is a strong predictor of house price, particularly in the higher-value segments of the market.

PDP for num_floors This plot shows a positive stepped relationship between the number of floors and the predicted house price. From 1.0 to 2.0 floors, the price increases modestly. A notable jump occurs between 2.0 and 2.5 floors, indicating that homes with more than two stories are associated with significantly higher values, likely due to architectural appeal or additional space.

After 3.0 floors, the predicted price plateaus, and the support becomes sparse. This means predictions beyond 3 floors are less reliable due to limited data.

In summary, homes gain most of their value from having more than one or two floors, but returns diminish beyond that.

4.3 PDP for sqft_living

This plot demonstrates a clear and strong positive relationship between living area (sqft_living) and predicted house price. Between 500 and 3000 sqft, the price increases sharply, showing strong marginal gains for added space.

Between 3000 and 4000 sqft, the curve flattens, indicating diminishing returns. However, after 7000 sqft, the price increases again and eventually levels off beyond 9000 sqft. This late rise could reflect the effect of luxury homes with premium pricing.

Again, the support ticks show that very large homes are less common, and predictions in those extremes should be interpreted carefully.

4.4 Final Conclusions

The PDP analysis of the random forest regression model reveals several important insights about how specific home features influence the predicted house price:

Living area (sqft_living) and number of bathrooms are the strongest predictors, showing clear positive and often nonlinear relationships with price.

Number of bedrooms exhibits a non-intuitive pattern, where more bedrooms can actually reduce the predicted price, potentially due to correlated factors like layout inefficiencies or neighborhood.

Number of floors has a positive but stepped effect, with the most significant price gains seen when moving from one to more than two floors.

In all cases, regions with low support (few observations) should be interpreted cautiously, as the model may be extrapolating beyond the training data.

These insights are valuable for understanding model behavior and for making informed decisions in real estate analytics or automated pricing systems.