XAI: Model-agnostic methods

Report 5 - Model-agnostic: Partial Dependency Plot (PDP)

Report written by: Natalia Martínez Calabuig
Elena Orón García
Aitana Sebastià Espinosa

Year: 2023/2024

Table of contents

Table of figures

1. Summary

This analysis aims to provide insights using Partial Dependence Plots (PDP) to understand the influence of different features on the predictions made by machine learning models. PDPs offer a way to interpret model-agnostic explanations, which are valuable in various applications such as predicting bike rental counts and house prices.

2. Introduction

We will need the files *day.csv*, which contains daily counts of rented bicycles from the bicycle rental company Capital-Bikeshare in Washington D.C., along with weather and seasonal information. This dataset not only facilitates the prediction of bike rental counts but also opens avenues for event detection and anomaly identification.

We will also use the file *kc_house_data.csv*, which contains the prices of houses, based on features bedrooms, bathrooms, floors, etc.

For the analysis of predictions for both the number of bikes rented and house prices, we are going to use the Partial Dependence Plot (PDP), which is a model-agnostic method. This type of model involves separating the explanations from the machine learning model, which provides flexibility.

A Partial Dependence Plot (PDP) is an interpretative tool that shows the marginal effect of a feature on the predicted outcome of a previously fit model. It helps to understand the relationship (linear, monotonic or more complex) between one or more features (independent variables) and the target variable (dependent variable) in a machine learning model. In other words, it shows how the model's prediction changes based on the values of a specific feature while keeping other features constant. This allows for visualising the isolated effect of a feature on the prediction.

A PDP can be one-dimensional, showing the relationship between one feature and the target variable, or multidimensional, showing the joint relationship between two or more features and the target variable.

However, some limitations of this method are:

- PDPs can be misleading if there are strong interactions between features. In such cases, the dependencies shown may not accurately reflect the model's complexity.
- Additionally, since it is a global method, PDPs assume an average relationship, which may overly simplify the true nature of interactions in the model.

3. Rental bikes database

We start by analysing the database of rented bicycles. To do this, we fit a random forest model to predict the variable *cnt*, which is the number of bikes rented. Using PDP plots we can visualise the relationships that the model has learned.

In these graphs, the black marks on the x-axis represent the distribution of the variable represented, which shows the relevance of the regions for interpretation, so we must be extremely cautious when interpreting regions with little data.

3.1. One dimensional Partial Dependence Plot

Below, we analyse the influence of days since 2011, temperature, humidity and wind speed on the predicted bike counts.

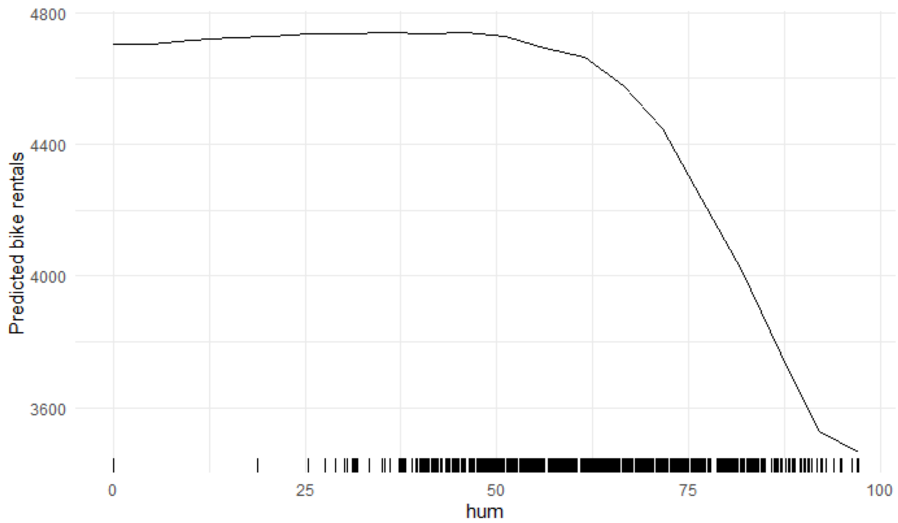

Figure 1. Partial dependence plot for Humidity.

Figure 1 shows the PDP for the variable humidity. In the low to moderate humidity range (0% to 60%), the number of rented bicycles remains relatively high and stable at around 4700 units. The curve is almost flat, indicating that humidity in this range does not have a significant impact on the prediction of the number of rented bicycles. As humidity increases beyond 60%, there is a noticeable drop in the number of rented bicycles. This drop is more pronounced as the humidity approaches 100%, down to approximately 3400 units. Thus, it appears that user comfort starts to be significantly affected when the humidity rises above the 60% humidity threshold.

It can be seen that there is a higher density of data in the 45% to 80% humidity range, so we can be more confident in the interpretations in this range. Predictions for humidity below 25% are not reliable, as no data is available.

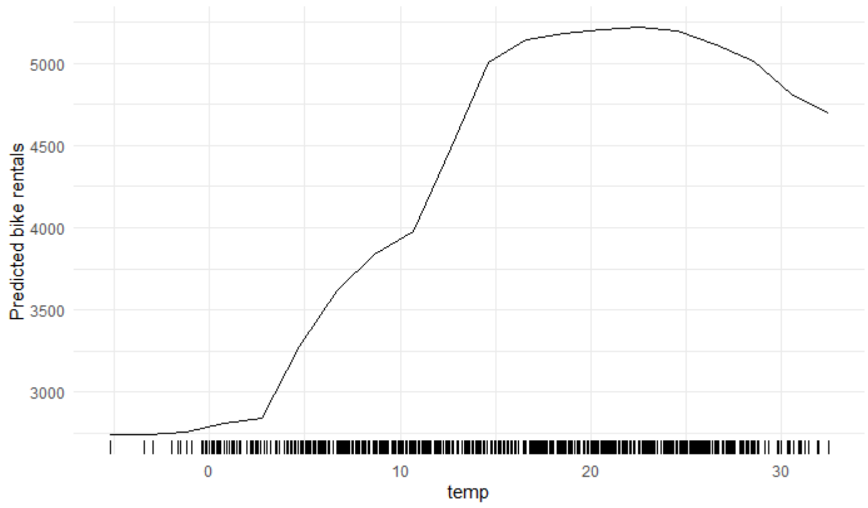

Figure 2. Partial dependence plot for Temperature.

Figure 2 shows the PDP for the temperature variable. First, we see that in the low temperature region (-0°C to 5°C), the number of bicycles rented is relatively low. This suggests that in cold climates, the demand for bicycles is lower (about 3300 bicycles).

As the temperature rises to 15°C, the demand for bicycles increases, first gradually and then sharply. After that, the prediction remains almost constant when the temperature is between 15ºC and 25ºC. The maximum number of bicycle rentals is predicted at around 20°C to 25°C. This is the point where the curve reaches its maximum height, exceeding 5000 rentals.

However, from 25°C onwards, the demand starts to decrease. This indicates that, although warm temperatures are generally favourable for bicycle rentals, such hot temperatures can deter people from renting bicycles.

Furthermore, we can be fairly confident in the interpretations, as most of the data concentrates on temperatures from 5°C to 28°C.

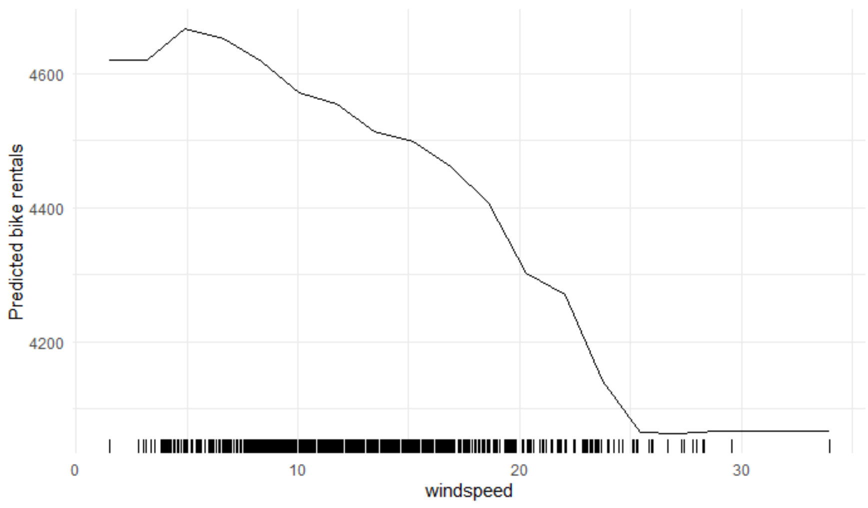

Figure 3. Partial dependence plot for Wind speed.

Figure 3 shows the PDP for the variable windspeed, which represents the wind speed in km/h. We see that in the range of 0 to 10 km/h, the number of rented bicycles remains relatively stable at around 4600 rentals. However, above 10 km/h, a gradual decrease in the number of rented bicycles is observed. And, between 20 and 30 km/h, the decrease in the number of rented bicycles becomes more pronounced, reaching about 4100 rentals. From 30 km/h onwards, the number of rented bicycles remains constant at the lowest point observed (approx. 4100 rentals).

That is, as wind speed increases, the number of rentals decreases, suggesting that moderate to high wind speeds deter users from renting bikes. Very high speeds seem to stabilise at a low level of demand. In contrast, low wind speeds do not have a significant impact on bicycle rentals.

However, the density of observations indicates that most of the data is concentrated at wind speeds below 20 km/h, with predictions for the range above 20 km/h being less reliable, as the number of data is reduced. Also, there are no instances when the windspeed is below 2 km/h, so we cannot assess the prediction of bikes in that range.

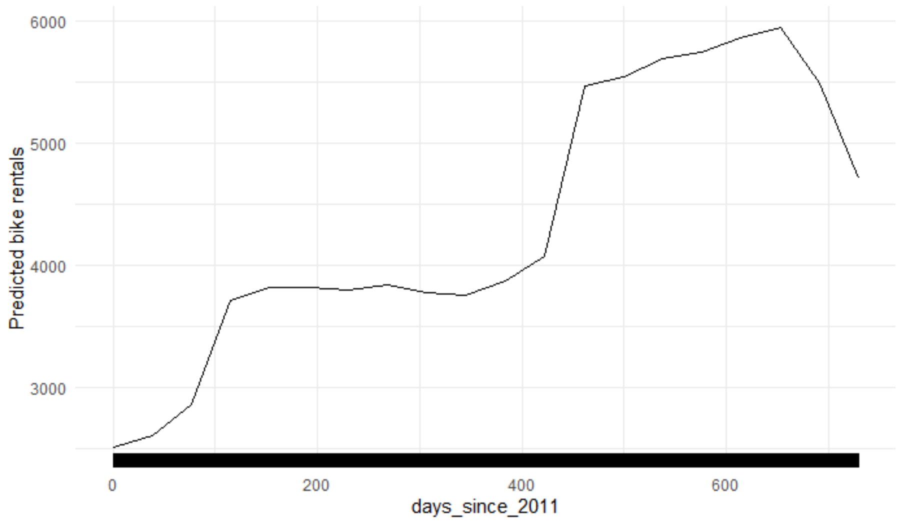

Figure 4. Partial dependence plot for Days since 1/1/2011.

Figure 4 shows the PDP for the variable days_since_2011. It represents the time in days since the first day of 2011, covering just over 600 days, which roughly covers the years 2011 and 2012. It should be noted that the density of observations indicates that data is available for almost every day in this time range, so the interpretation we make of the relationship is relevant.

At the beginning, from 0 to 100 days, there is a rapid increase in the number of rented bikes. This could correspond to the end of winter and the beginning of spring in 2011, when the weather improves and more people start renting bikes.

Then, between days 100 and 400, the number of rented bikes remains relatively stable around 4000. This could correspond to the summer and autumn months of 2011, where weather conditions are favourable for bicycle rentals.

From day 400 onwards, there is a sharp increase in the number of rented bikes, reaching 5500 rentals. After day 460, rentals continue to increase, although less than before, reaching a peak of around 6000 rentals.

This increase could be associated with improvements in the bike rental service or a popularisation of the company or urban cycling in general.

Finally, from day 650 onwards, there is a noticeable decrease in the number of bikes rented. This decrease could be due to the onset of winter 2012, where less favourable weather conditions reduce demand.

In summary, the PDP shows peaks in bicycle rentals during the warmer periods and declines in the colder months, which could indicate seasonality.

3.2. Bidimensional Partial Dependence Plot

We can also obtain the relationship between two predictors by means of a heatmap. In Figure 5 we have generated a 2D Partial Dependence Plot with humidity and temperature to predict the number of rented bicycles as a function of these parameters. We also show the density distribution of both input features.

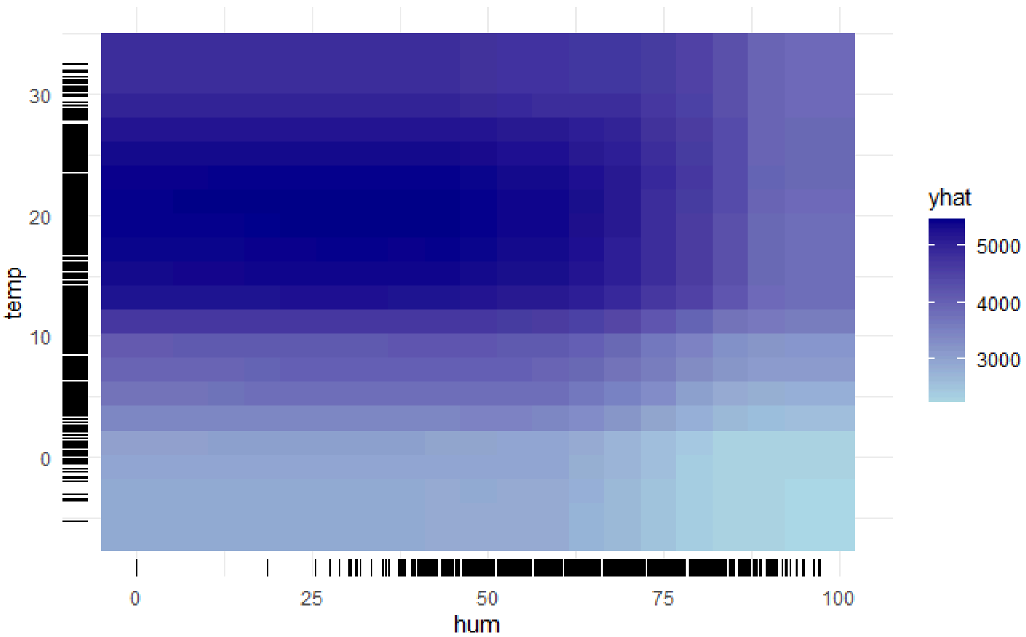

Figure 5. Bidimensional partial dependence plot for Humidity and Temperature.

On the one hand, on the X-axis of Figure 5, humidity varies from 0% to 100% and there is a higher density of observations in humidity ranges between 30% and 90%. On the other hand, on the y-axis, temperature varies from approximately -10 to 35 degrees Celsius, but there is a higher density of observations in temperature ranges between 0 and 30 degrees Celsius. And, in terms of colours, deep blue indicates a higher number of rented bicycles while light blue indicates a lower number of rented bicycles.

We see that the darker areas, where more bicycles are rented, are in the higher temperature ranges (15ºC to 30ºC). Within this high temperature range, there seems to be no clear trend in humidity, but the range between 40% and 75% humidity seems to be favourable for bicycle rentals.

Likewise, the clearest areas, where fewer bicycles are rented, are observed in the lower temperature ranges (below 10 degrees Celsius) and at very high humidity levels (above 80%).

In general, it appears that the influence of humidity is less pronounced than that of temperature. While temperature shows a clear relationship with bicycle rentals (higher temperatures lead to more rentals, up to a certain threshold), humidity has a more moderate effect.

In summary, the graph suggests that bicycle rentals are highest when temperatures are moderate to high (20°C to 30°C), while humidity does not have such a clear impact, although it seems that extremely high humidity levels might be the least favourable for rentals.

4. House prices database

Now, we apply the previous concepts to predict the price of a house from the database kc_house_data.csv. In this case, we again use a random forest approximation for the prediction based on the features bedrooms, bathrooms, sqft_living, sqft_lot, floors and yr_built. Due to the size, we extracted a set of random samples (25%) from the BBDD before fitting the random forest model. Then, we use the partial dependence plot to visualise the relationships the model learned.

4.1. One dimensional Partial Dependence Plot

Below, we analyse the influence of bedrooms, bathrooms, sqft_living, sqft_lot, floors and yr_built on the predicted price.

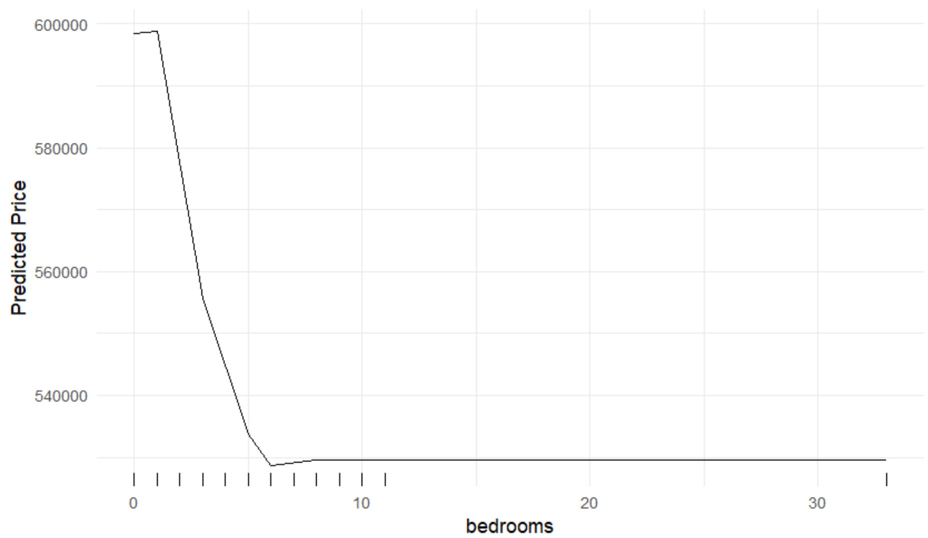

Figure 6. Partial dependence plot for number of bedrooms.

In Figure 6, we observe the bedrooms variable. We can state the following:

While the x-axis represents the number of bedrooms in a house, the y-axis represents the predicted house price and shows how the predicted price changes as the number of bedrooms increases. The values of the bedrooms are discrete, ranging from 0 to over 30 bedrooms, though most realistic values are clustered at the lower end of this range.

At 0 bedrooms, the predicted price is quite high, but there is a steep drop as the number of bedrooms increases to around 1-2 bedrooms (from $580,000 to $530,000). This drop might indicate an anomaly or an artificial dataset, as houses typically have at least one bedroom, and the model may be extrapolating unusually for these unrealistic cases.

As the number of bedrooms increases from around 3 to 10, the predicted price recovers and stabilises. The predicted price remains relatively flat for houses with more than 10 bedrooms, indicating that adding more bedrooms beyond this point does not significantly affect the predicted house price. It seems that the market values these houses similarly, regardless of additional bedrooms.

The stability within this range may suggest that such properties are rare and do not follow common trends in the residential market. They could belong to a specialised segment of the real estate market, such as commercial properties, multifamily residences, or large mansions where the number of rooms is not the primary determinant of price.

Note that the graph displays a higher density of data for houses with fewer than 10 rooms, indicating that most properties in the dataset have this number of bedrooms. Therefore, predictions within this range are more reliable. For properties with more than 10 bedrooms, the lower data density suggests that these are less common, and predictions within this range may be less accurate. It might reflect issues with the model's extrapolation or a lack of data points for these extreme cases. This behaviour is unusual and should be interpreted with caution. It might be useful to check the data distribution, for example.

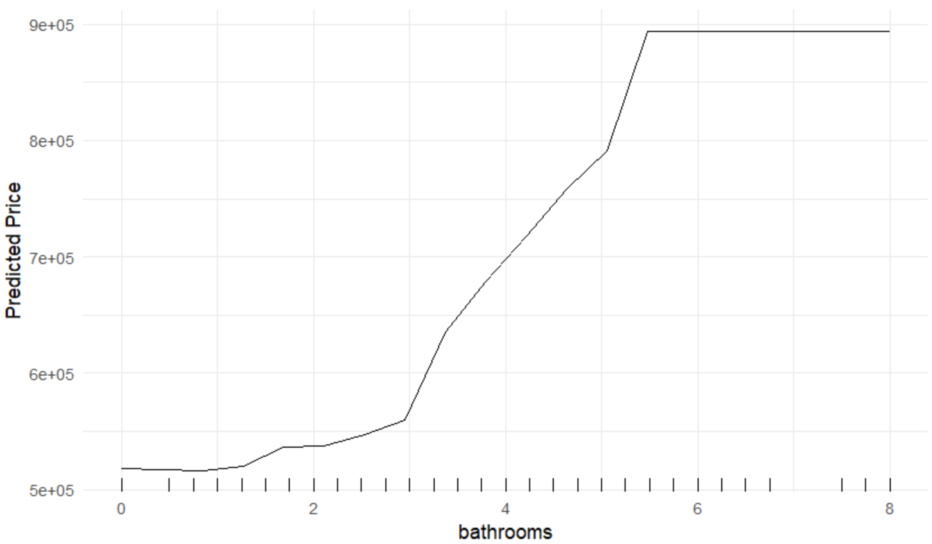

Figure 7. Partial dependence plot for number of bathrooms.

Figure 7 shows the PDP for the variable bathrooms, which is the number of bathrooms in the house.

In the range of 0 to 3 bathrooms, the predicted house price slowly increases from approximately \$500,000 to \$600,000. This gradual increase suggests that adding one or two additional bathrooms has a moderate positive impact on the price of the house.

Between 3 and 4 bathrooms, the predicted price continues to increase, reaching around \$700,000. The rate of increase is most notable in this range, indicating that homes with 3 to 4 bathrooms are considerably more valuable than those with only 2 bathrooms.

The predicted price experiences a steeper increase in this 4-6 bath range, rising from approximately \$700,000 to \$900,000. This range shows a strong positive correlation between the number of bathrooms and the price of the house. This pattern suggests that from 4 bathrooms upwards, properties are likely to belong to a more exclusive and higher value market segment, where additional bathrooms are a luxury feature.

From 6 bathrooms onwards, the predicted price stabilises and reaches a ceiling close to \$900,000. This behaviour suggests that, after a certain point, adding more bathrooms does not contribute significantly to the price increase.

The data density appears to be higher for houses with up to 4 bathrooms, suggesting that these are more common in the dataset. Therefore, predictions in this range are more reliable. For houses with more than 4 bathrooms, the lower data density may reduce the reliability of predictions in these ranges.

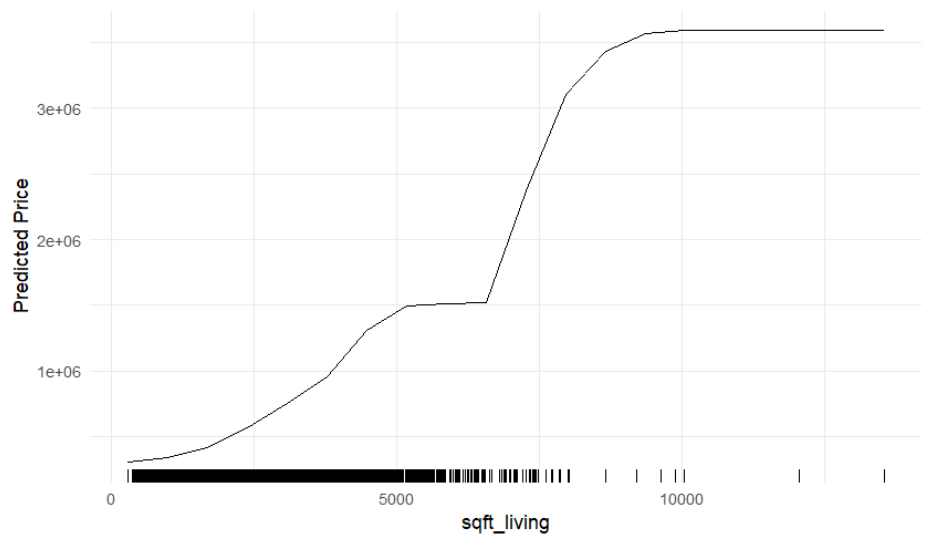

Figure 8. Partial dependence plot for SQFT living.

Figure 8 shows a one-dimensional Partial Dependence Plot illustrating the relationship between the variable sqft_living (square feet of living area) and the predicted price of a house.

In the range 0 to 5000 sqft, the predicted house price increases gradually as the living area increases, from \$500,000 to \$1,500,00. This indicates that houses with larger sqft tend to have higher prices. After that, it remains stable up to about 6500 sqft.

As the living area continues to increase beyond 6500 sqft, the predicted price increases more steeply, up to over $3,500,000. This is the range where the relationship between the size of the living area and the price of the house is strongest, indicating that increases in living area have a significant impact on price.

When the sqft approaches 10,000, the price continues to increase but at a slower rate, showing a stabilising curve. And, when the sqft is greater than 10,000, the predicted price remains constant, i.e., once this point is reached, in extremely large houses, other factors (such as location or design) may have a greater influence on the price.

It is worth noting that we observe a higher density of data in the 0 to 5000 sqft range, so the interpretation for larger houses would not be reliable, as we do not have enough data.

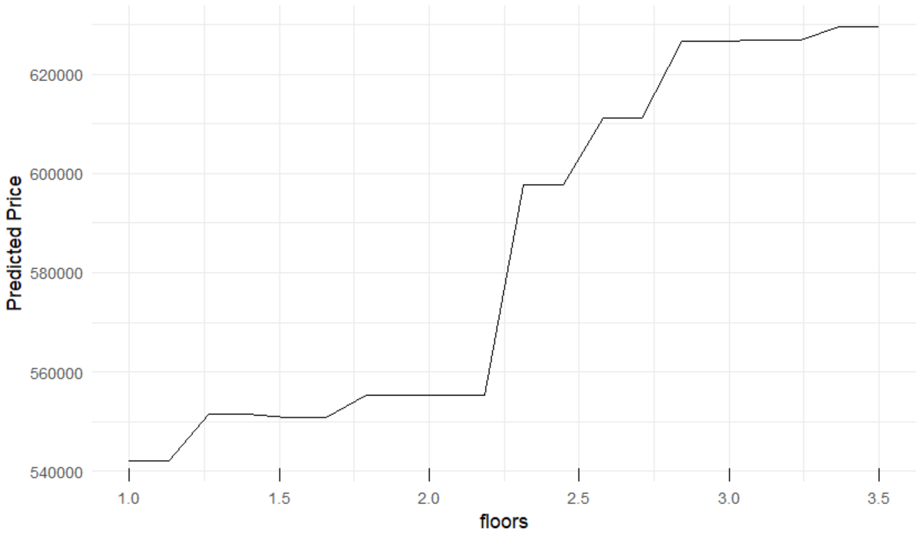

Figure 10. Partial dependence plot for number of floors.

In Figure 10 we can see on the x-axis the values of floors, which goes from 1 to 3.5, while the y-axis represents the predicted house price. The line shows the average predicted price for different values of the floors feature while keeping all other features constant.

As the number of floors increases from 1.0 to 1.5, there is a slight increase in the predicted price and it seems to remain constant until reaching the number 2 of floors. This might indicate that slightly more floors could initially be associated with a lower price, potentially due to other correlated factors not shown in this plot.

Starting from 2.0 floors, there is a noticeable increase in the predicted price, which continues with additional floor until reaching its peak in number of floors of 3. A notable upward trend in predicted price starts at around 2.0 floors, indicating that houses with more than 2 floors are predicted to be more expensive. This trend continues significantly up to 3.0 floors.

After 3.0 floors, the increase in predicted price plateaus, showing little to no change as the number of floors approaches 3.5. This suggests that adding more floors beyond 3.0 does not significantly increase the predicted price. The market may perceive houses with more than three floors similarly in terms of value.

Note that these decimal values in the context of the PDP are more of a visual and analytical tool, and do not directly reflect real world data. Although in reality there is no half-floor, this interpolation allows to visualise a continuous trend and to better understand how the price might change with small changes in the number of floors. It helps to get a clearer picture of the behaviour of the model and to identify points where a change in the number of floors has a greater impact on the price.

To sum up, the PDP in Figure 9 indicates that the number of floors has a varying impact on predicted house prices. Initially, more floors seem to have a neutral to slightly negative effect, but after a certain point (around 2.0 floors), additional floors lead to higher predicted prices. Beyond 3.0 floors, the effect stabilises, suggesting diminishing returns on house prices for adding more floors. This understanding helps in assessing how the model values the floors feature in the context of predicting house prices.

4.2. Bidimensional Partial Dependence Plot

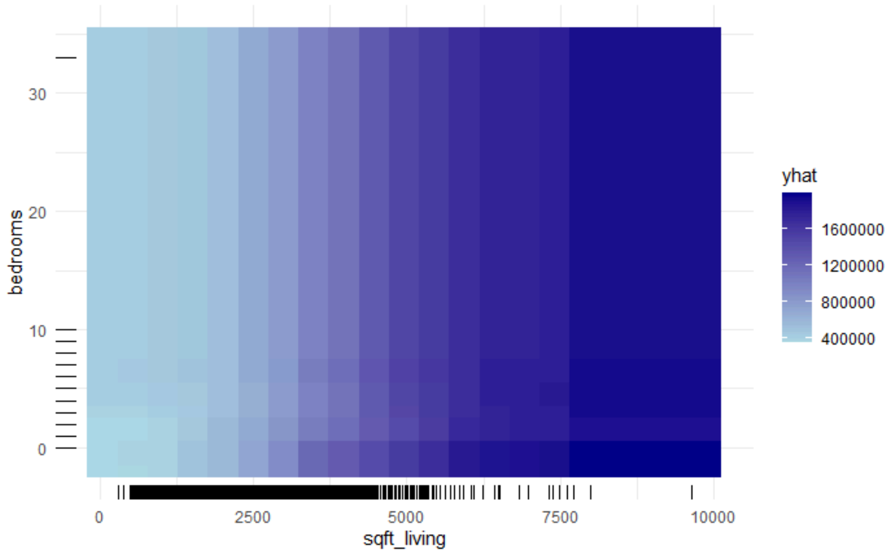

Figure 11. Bidimensional partial dependence plot for sqft_living and bedrooms.

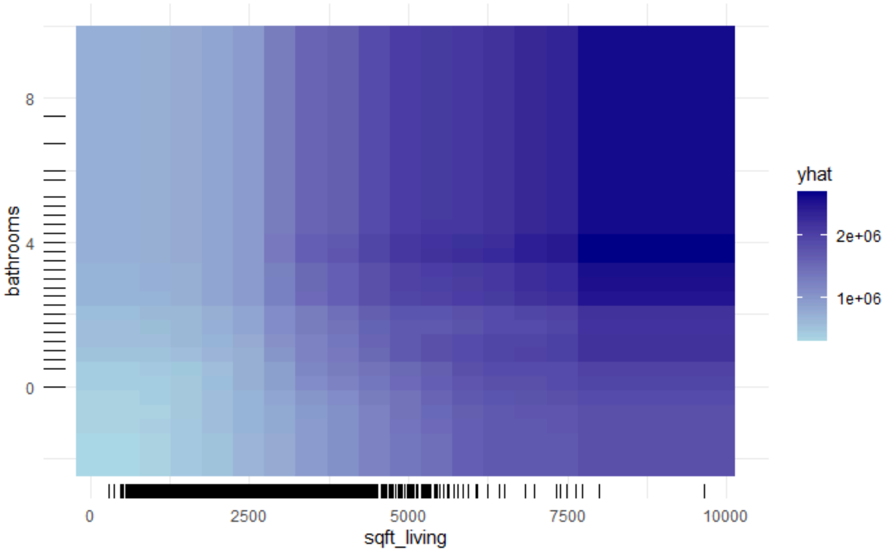

Figure 12. Bidimensional partial dependence plot for sqft_living and bathrooms.

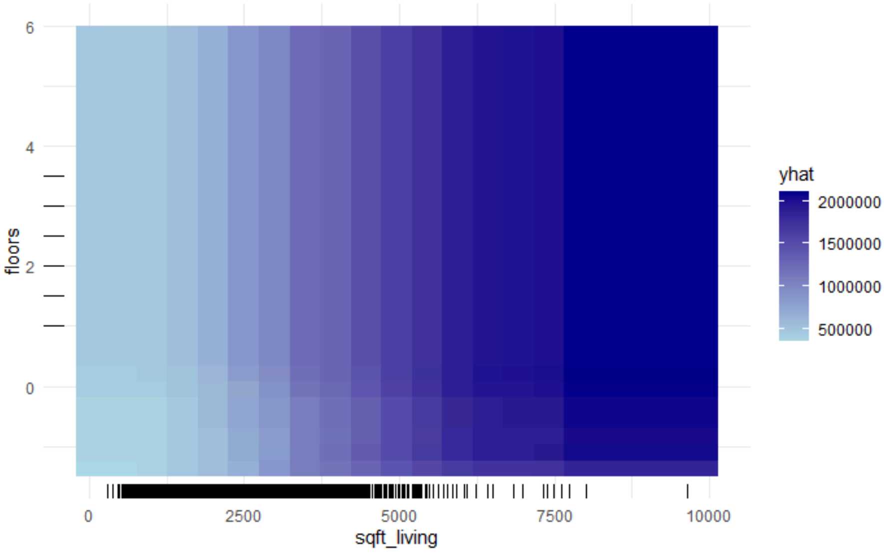

Figure 13. Bidimensional partial dependence plot for sqft_living and floors.

Figures 11, 12 and 13 are bidimensional partial dependence plots showing the joint relationship between two characteristics. In all three graphs, the x-axis represents the amount of sqft living of the house, varying from 0 to 10,000. As they all show a similar pattern, we can interpret:

The lighter areas (lower predicted prices) are at the bottom left of the graph, where there are few sqft living and few bedrooms/bathrooms/floors. The darker areas (higher predicted prices) are in the upper right, indicating that houses with many sqft living and many bedrooms/bathrooms/floors have higher predicted prices.

It should be noted that when the number of bedrooms/bathrooms/floors increases, the predicted price increases very slightly compared to the increase in sqft living. In addition, one has to be careful with interpretations for sqft living above 5,000 as there is little data.

5. Conclusion

In conclusion, for the bicycle rental business, it has been identified that optimal climatic conditions to promote rental include moderate temperatures (20°C to 25°C) and humidity not exceeding 80%. Therefore, it is recommended to focus promotion strategies during these favourable periods and consider offering incentives or adjusting operations during less favourable conditions, such as extreme temperatures, high humidity or strong winds.

In terms of house prices, a strong positive correlation has been observed with the number of bathrooms and the size of the living area. And, although a larger number of bedrooms may initially decrease the price, this effect stabilises later on. In addition, homes with more floors generally tend to have higher prices.

Finally, Partial Dependence Plots (PDPs) are valuable tools for model interpretation, as they provide information on how individual characteristics and their interactions influence predictions. This is essential for understanding model behaviour and making informed decisions based on the analysis.