XAI: Model-agnostic methods

Vlad Galan, Angel López, Guillermo Lladró

1. Introduction.

In the realm of explainable artificial intelligence (XAI), model-agnostic methods provide crucial insights into the behaviours of predictive models, particularly those considered as black boxes. This exercise delves into the utilization of Partial Dependency Plots (PDP), a model-agnostic technique that reveals the relationship and impact of predictor variables on the outcome predicted by a model. Throughout this task, it will be employed Partial Dependency Plots to understand the influences of specific features in diverse contexts: predicting the number of bike rentals and house prices.

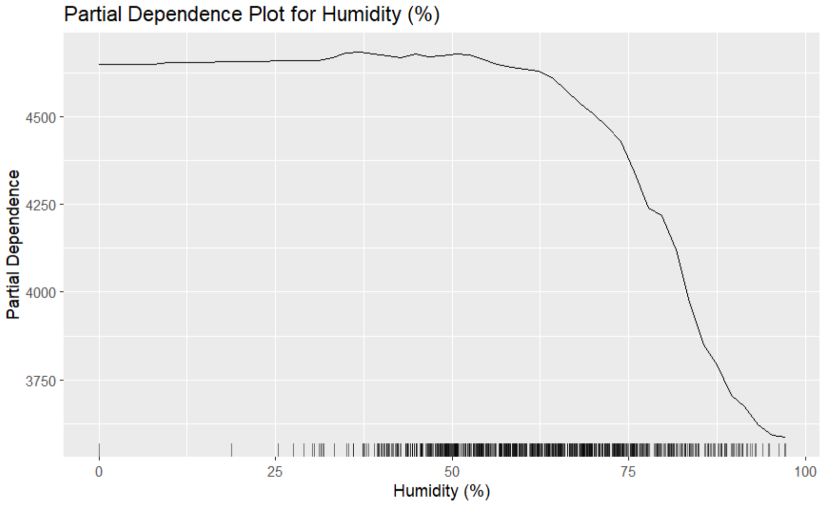
2. Analysis.

The first part of the practice focuses on the application of one-dimensional PDPs to a random forest model trained to predict bike rentals. Variables such as days since 2011, temperature, humidity, and wind speed are analyzed to discern their effects on bike rental counts. The second part of the exercise extends the PDP analysis to two dimensions, examining the interplay between humidity and temperature on bike rental predictions.

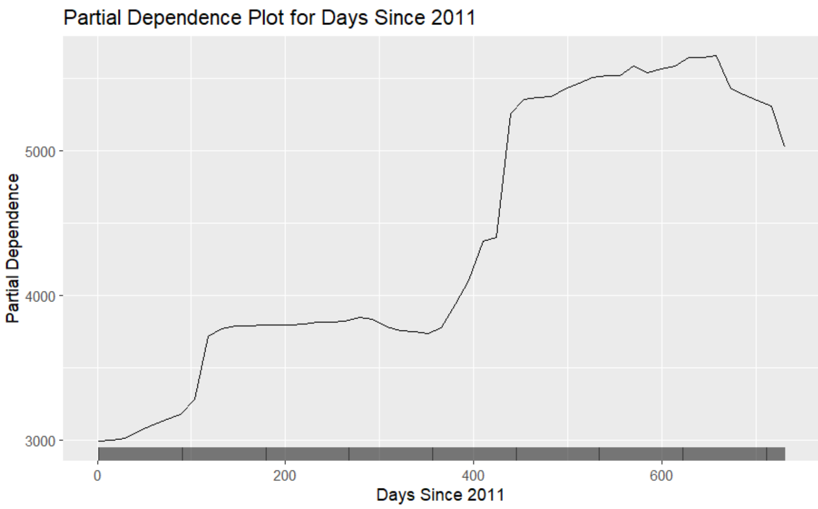
The third and final part of our exploration applies PDP to the real estate sector, using a random forest model to predict house prices from features such as the number of bedrooms, bathrooms, square footage of living space, and number of floors. By using PDPs, we aim to unpack the influence of these features on housing prices, providing a deeper understanding of market dynamics.

2.1. One dimensional Partial Dependence Plot.

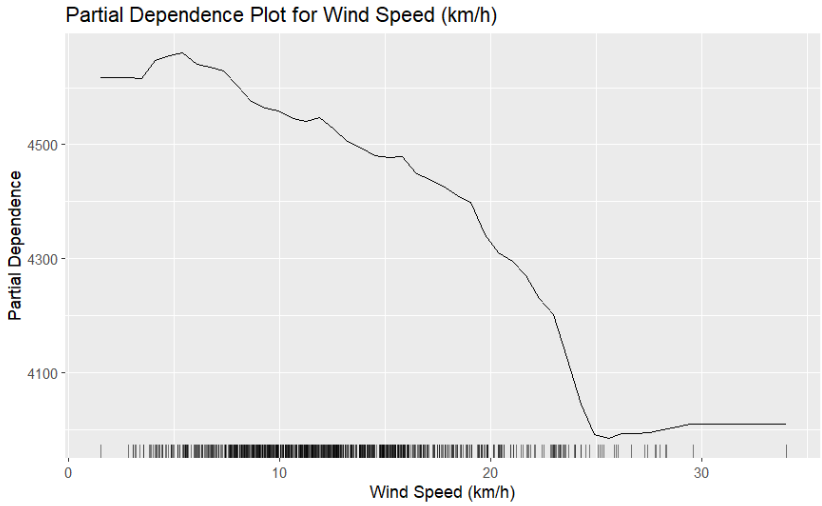
In this section it will be analysed how some of the variables of the random forest model we have used affect the prediction of the number of bicycles rented. This analysis will be carried out using PDP (partial dependence plot) graphs, a type of graphs that are characterised by their versatility since they can be obtained independently of the model we have used, including black box models. The features that are going on to be considered for the analysis are: ‘Humidity’, ‘days_since_2011’, ‘wind speed’ and ‘temperature’.



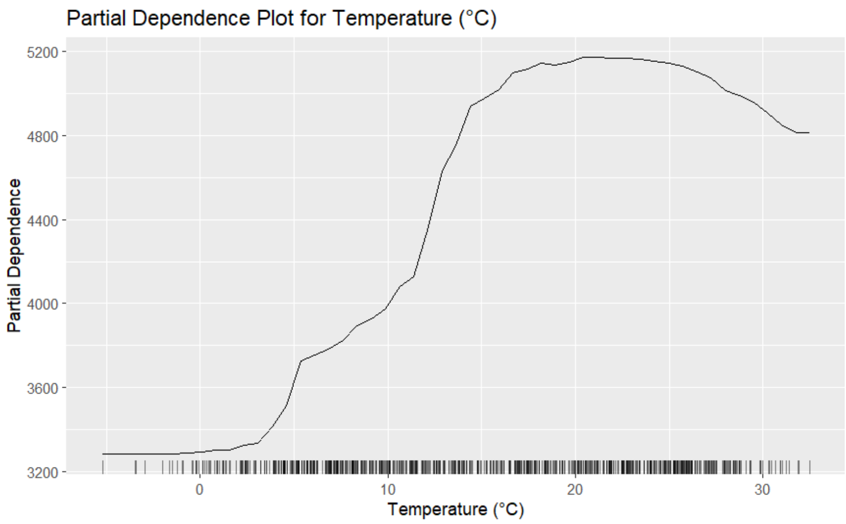
Firstly, we will analyse the PDP for the variable Humidity, we observe that the expected number of rented bicycles is higher than 4500 for humidity values between 0 and 60, however we cannot obtain precise conclusions on this entire range of values as the density of observations between a percentage of 0 and 25 per cent humidity is very low, from 25 to 60 we do find enough instances to take into account the results. From 50 percent humidity (at which point we still have around 4500 rented bikes), an increase in humidity means a decrease in the expected number of rentals, with an expected value of around 3600 for a humidity of 100. From this point onwards, the number of recorded values is very low, and it is no longer possible to infer results.



In the case of the variable 'days_since_2011' it seems that as the number of days increases the number of rentals also increases. Between 100 and 400 days elapsed the number of rentals remains stable at around 3900 bikes, but between 400 and 430 there is a very pronounced change that reaches more than 5000 predicted rentals, and this point is maintained until more than 600 days elapsed where we observe a small drop in the number of loans.



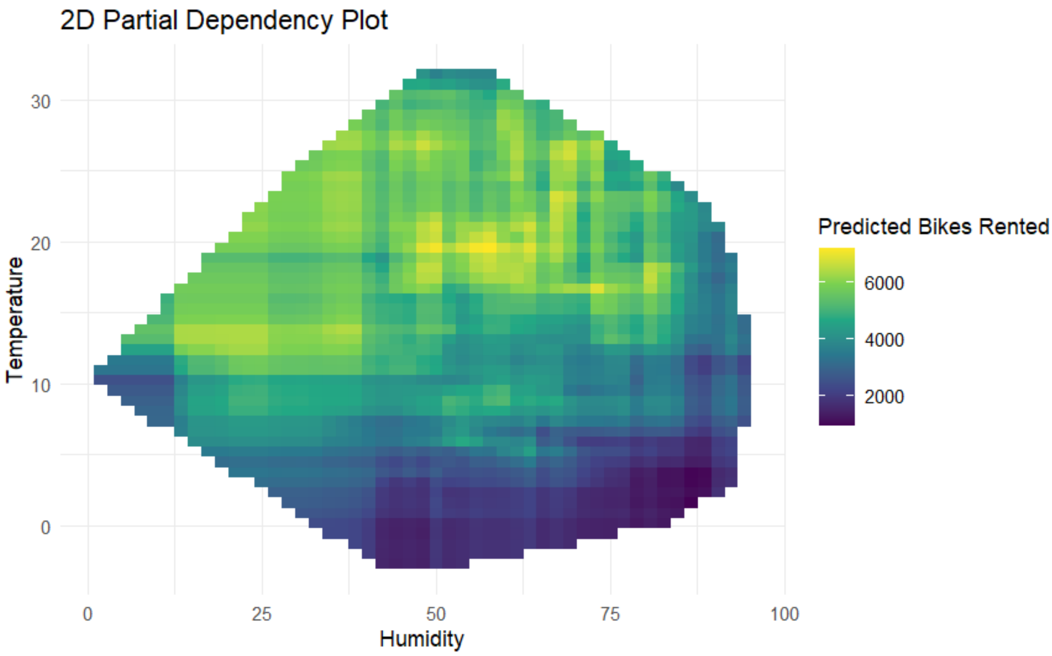
Looking now at the wind speed, it can be seen that as the wind speed increases the number of bicycles decreases from about 4600 rented without wind to less than 4100 at wind speeds of 25, where the decrease stabilises and the number stops decreasing even above the value of 30. Despite the above results, at both extremes of wind speed, i.e. temperatures near 0 or below and near 30 or above the conclusions are not applicable as the number of observations is very low and the results may be biased.



Finally, in the case of temperature, the predicted rents increase up to 15 degrees, starting from 3200 predicted at 5 degrees and reaching more than 5000, we do not take into account values around 0 degrees and below as we observe very little density of observations. Between 15 and 25 degrees the value stabilises at around 5200 rentals and from that point onwards the figure starts to decrease less sharply with a prediction of 4800 at 30 degrees. From this point onwards there are few observations again and the results could be biased by concordant situations.

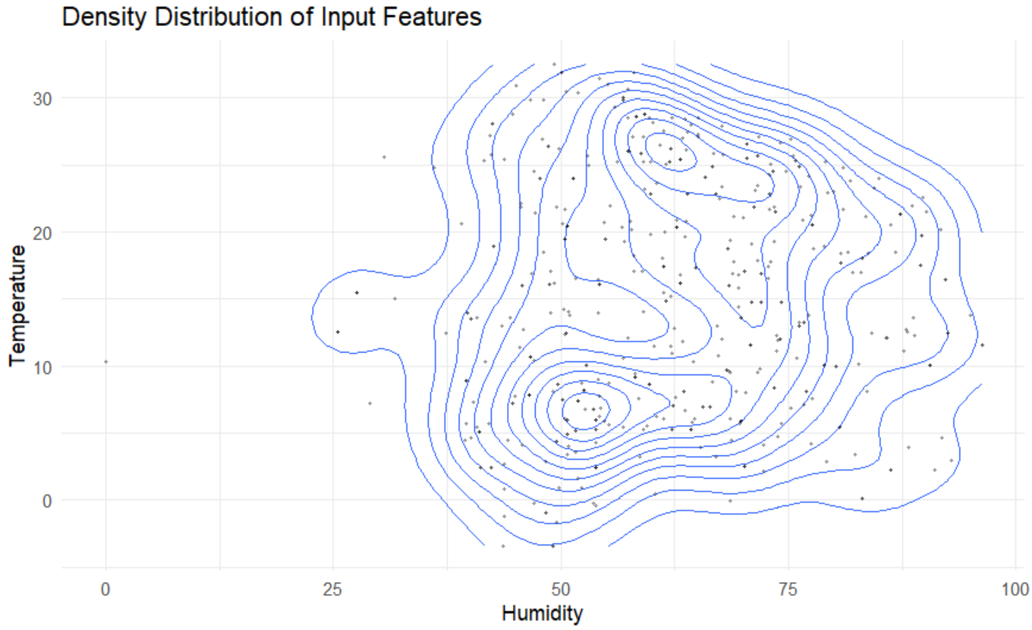
2.2. Bidimensional Partial Dependency Plot.

The next step is to observe how the number of rented bikes varies as a function of the humidity and temperature parameters. For this purpose, a 2D PDP (Partial Dependence Plot) will be constructed, which will show the marginal effect of both variables on the target variable (number of bikes rented).



It is clearly observed that for lower temperatures, at any humidity level, the number of rented bicycles is at a rather low level. On the other hand, for higher temperatures, especially in the range between 15 and 25ºC, and with humidity levels between 50 and 75%, a very favourable case for bicycle rentals is suggested. It is also relevant to note that, if there is a very high level of humidity, even if temperatures are moderate, the number of bikes rented will decline. This suggests that in good weather, people are more inclined to cycle.

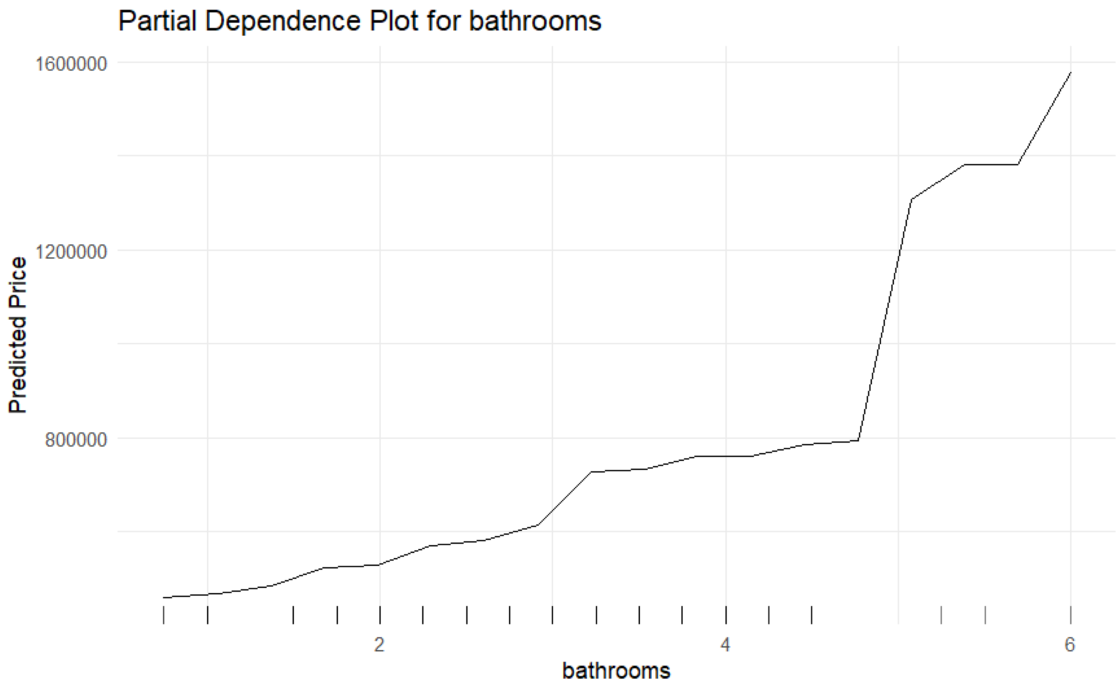
However, by constructing a density plot, it is possible to validate whether the patterns observed in the PDP are due to areas of high data density. If a strong pattern in the PDP coincides with high data density, it is more likely to be a reliable model output.



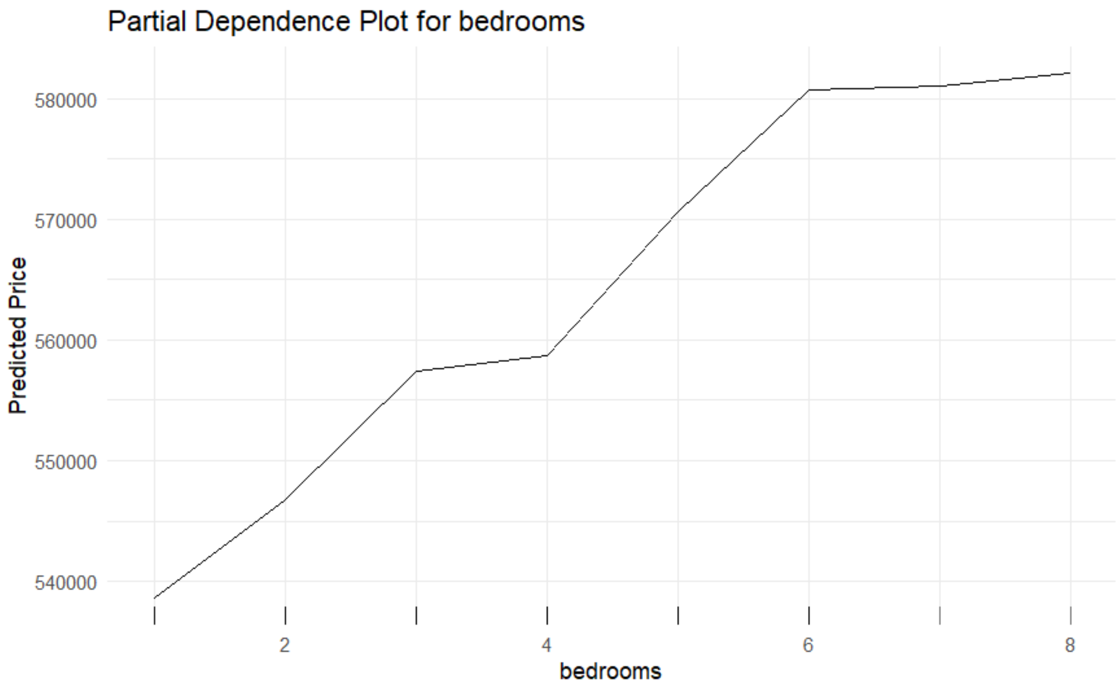
Thus, it is observed that for temperatures below 10 there is a high concentration of data, which can be interpreted as a reliable prediction by the model. On the other hand, when attention is focused on the most favourable situation for bicycle rental, it is found that such combinations of temperatures and humidity are less common in the data, which could imply that, although these are the times with the most bicycles used, they are situations that do not occur so often. Therefore, such areas with more separated contour lines result in less reliable predictions.

2.3. PDP to explain the price of a house.

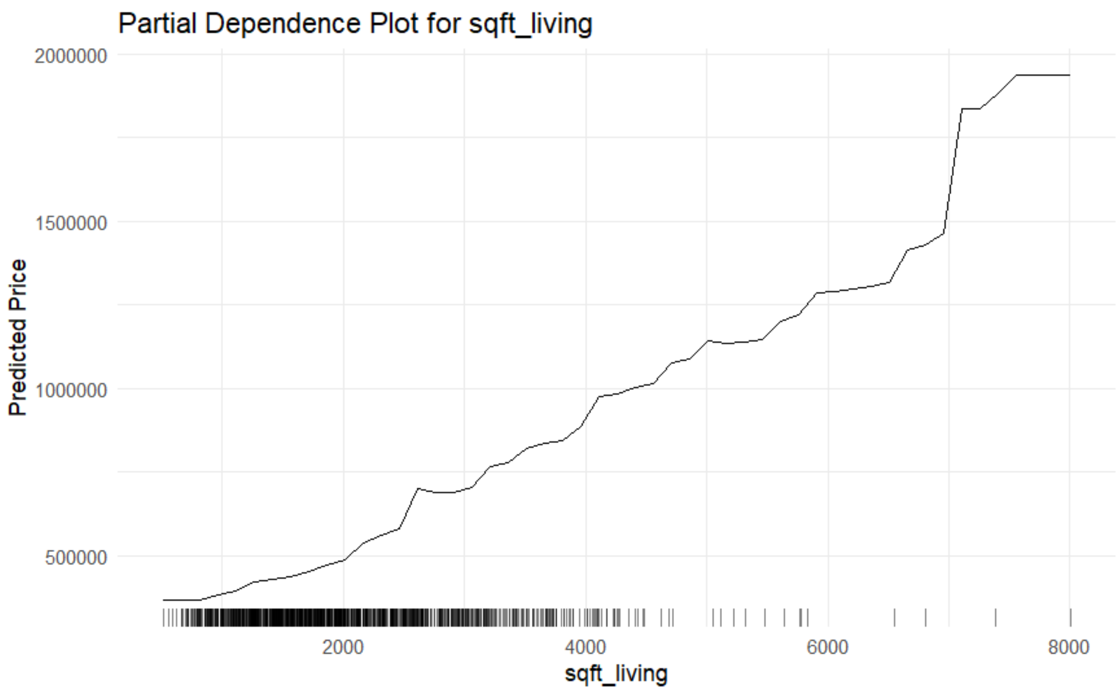
Now, just as was done in the first section, the influence of various house characteristics on the housing price will be analyzed. These characteristics include the number of bathrooms, bedrooms, square footage of living space, and the number of floors. For this analysis, PDP (Partial Dependence Plot) will again be used to show the marginal effect of a feature on the predicted outcome of a previously fit model.



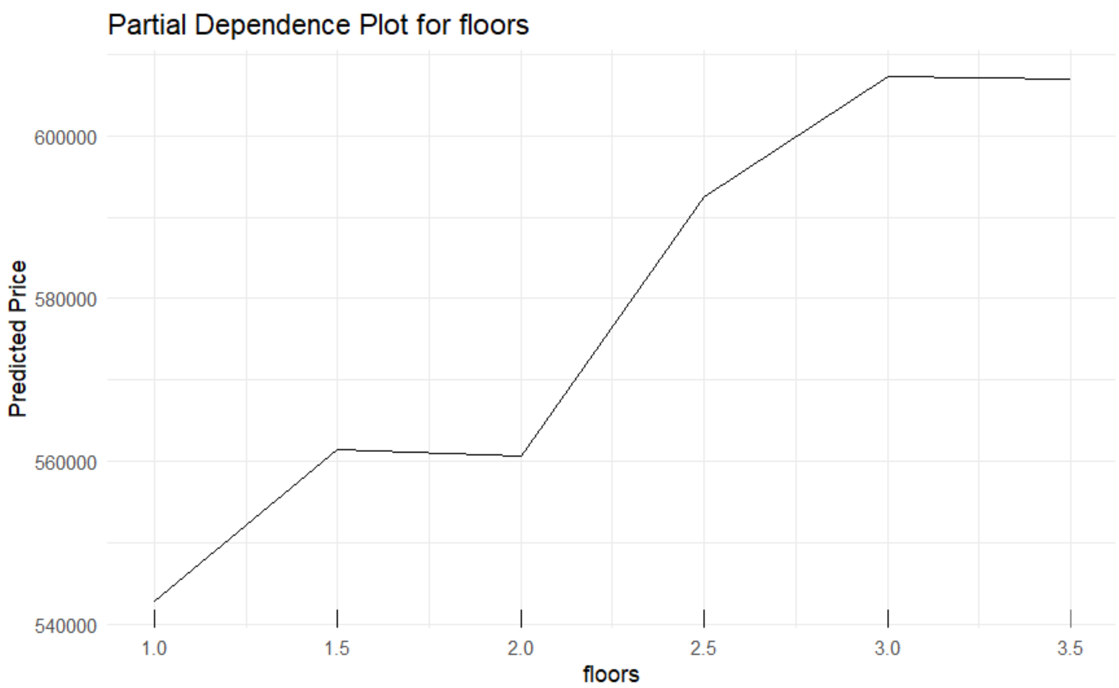
First, the influence of the number of bathrooms in the house was analyzed. As expected, the more bathrooms there are, the higher the expected price. Notably, there is a significant increase in price starting from the fifth bathroom. However, this peak should not be considered relevant due to the low sample density between 5 and 6 bathrooms.



For bedrooms, a similar pattern is observed. It is evident that the more bedrooms a house has, the higher the expected price. However, there are two ranges where the expected price seems to remain stable. The difference between 6 and 8 bedrooms does not follow the upward trend, but, as before, the number of samples is not significant enough to draw conclusions from this data. Regarding houses with 4 bedrooms, the expected price remains nearly the same as with 3 bedrooms. This could be because PDP measures the effect of variables without considering how the others fluctuate, so a house with 4 bedrooms and 1 bathroom might not have a considerable price difference compared to a house with 3 bedrooms and 1 bathroom.



When looking at the square footage, it again aligns with the idea that larger living spaces lead to higher property prices. However, drawing conclusions for properties with over 5000 square feet is not advisable due to the scarcity of samples at these values.



Lastly, the number of floors also positively affects house prices. Interestingly, this upward trend breaks at 2 floors. Market needs, such as a preference for single-story homes due to mobility or convenience, may explain this situation. Beyond 3 floors, valuable conclusions cannot be inferred due to the minimal number of examples in these cases.

3. Conclusions.

During the first part, it has been highlighted the application of one-dimensional Partial Dependence Plots to a random forest model predicting bike rentals. Through detailed analysis, it was observed significant influences of factors like humidity, days since 2011, wind speed, and temperature on rental counts. Notably, the analysis demonstrated that lower humidity levels correlate with higher bike rentals, which tend to decrease as humidity increases. Similarly, a positive correlation was identified between the elapsed days since 2011 and rental counts, reflecting perhaps an increasing trend in bike usage over time or improvements in data collection methods. The insights from wind speed and temperature further underscore the nuanced relationships these environmental factors have with bike rental behavior, providing actionable insights for city planners and bike rental companies to optimize their services according to prevailing weather conditions.

For the second part a two-dimensional Partial Dependence Plot has been used in order to observe how different combinations of temperature and humidity affect bicycle rentals. The combined use of this PDP with the density plot has allowed a deeper understanding of the dynamics between climate and cycling. It is observed that when there are areas with high data density, the model predictions are robust, but this is not the case when interpreting the results in areas with less data. These insights can guide strategic decisions to improve service and increase bicycle rental in the city.

Finally, the influence of house features on pricing was examined using Partial Dependence Plots again, offering a granular look at how bedrooms, bathrooms, square footage, and floors impact house prices. The analysis reaffirmed the intuitive assumption that more bathrooms and larger living spaces generally lead to higher property values. However, the PDPs also unveiled less obvious insights, such as the diminishing returns on adding more than two floors, and the complex interplay between the number of bedrooms and other house features like bathrooms, which may not always lead to straightforward increases in house prices, although it may be caused by the data sample chosen. These findings are critical for both sellers and buyers in the real estate market, as they highlight the importance of balancing various house attributes to maximize property value.