Content table

[1. One dimensional Partial Dependence Plot 2](#_Toc134028704)

[2.- Bidimensional Partial Dependency Plot 2](#_Toc134028705)

[3.- PDP to explain the price of a house. 3](#_Toc134028706)

## 1. One dimensional Partial Dependence Plot

The partial dependence plot shows the marginal effect of a feature on the predicted outcome of a previously fit model. In our case, we build a Random Forest model to predict the number of bikes rented using the features workingday, holiday, spring, summer, fall, misty, rain, den\_temp, den\_hum, den\_windspeed and days since 2011.

After building the model, we can use the partial() function to learn and then visualize the relationships the model learned. Specifically, we are going to study the influence of the number of days since 2011, the temperature (den\_temp), the humidity (den\_hum) and the wind speed. To do so, we build the following graph:

Gráfico, Gráfico de líneas

Descripción generada automáticamente

Thanks to this figure we can compare the influence of each of this four variables on the predicted bike counts.

For instance, if we look at the top left graph we can see that the number of bikes rented increases with the temperature until the temperature is around 15. Then it remains constant until the temperature is above 30 degrees and it starts decreasing.

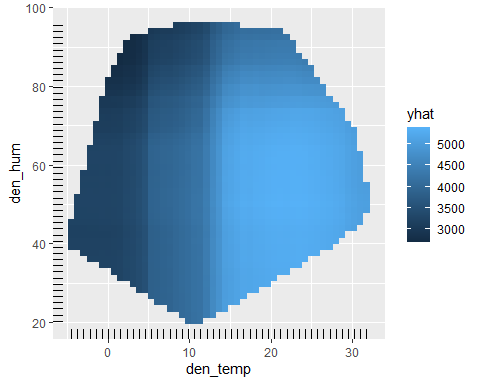
If we look at the top right, we can see that the humidity has almost no impact on the bike count as the number of bikes is constant except when the humidity is over 75, that makes it decrease.

The wind speed follows a similar pattern but, in this case, the number of bikes rented decreases gradually until the wind speed is above 20 when it remains constant.

Finally, we can study the influence of the number of days since 2011. In the correspondent graph we can see that the number of bikes rented increases with the number of days.

## 2.- Bidimensional Partial Dependency Plot

For this section, we want to analyze the influence in the prediction of the number of bikes rented of both humidity and temperature. To do so, we have built a partial dependency plot that includes these two features.



To interpret this plot it is important to take into account the plots generated in the previous section. This means that at first, we could get an idea of what the plot would look like in case there was no interaction between the variables. That is, in the previous exercise we have observed in the plot of the variable "den\_temp" that the value of the variable explained (cnt, which is the number of bicycles rented) increases as the value of the variable "den\_temp" increases. It is also observed in the graph of the variable "den\_hum" that the value of the explained variable decreases as the value of the variable "den\_hum" increases.

In short, what we could expect in advance from the 2D partial dependency plot, in case there was no interaction between the variables "den\_temp" and "den\_hum", is that the highest values of the explained variable are in the area where the values of the variable that measures humidity are low and the values of the variable that measures temperature are high. We would also expect the lowest values of the explained variable to be on the opposite side of the plot (where humidity is high and temperature is low).

As we observe that, indeed, the plot has the shape we expected, we can conclude that if there is some kind of interaction between the variables "den\_temp" and "den\_hum", we are unable to detect it through this dependency plot.

## 3.- PDP to explain the price of a house.

*EXERCISE:* Apply the previous concepts to predict the price of a house from the database kc\_house\_data.csv. In this case, use again a random forest approximation for the prediction based on the features bedrooms, bathrooms, sqft\_living, sqft\_lot, floors and yr\_built.

Use the partial dependence plot to visualize the relationships the model learned.

*BE CAREFUL:* due to the size, extract a set of random samples from the BBDD before generating the data for the Partial Dependency Plot.

*QUESTION:* Analyse the influence of bedrooms, bathrooms, sqft\_living and floors on the predicted price.

Interfaz de usuario gráfica, Gráfico

Descripción generada automáticamente