UNIVERSIDAD POLITÉCNICA DE VALENCIA

XAI 3: Model-agnostic methods

Group members:

Hugo López Juan

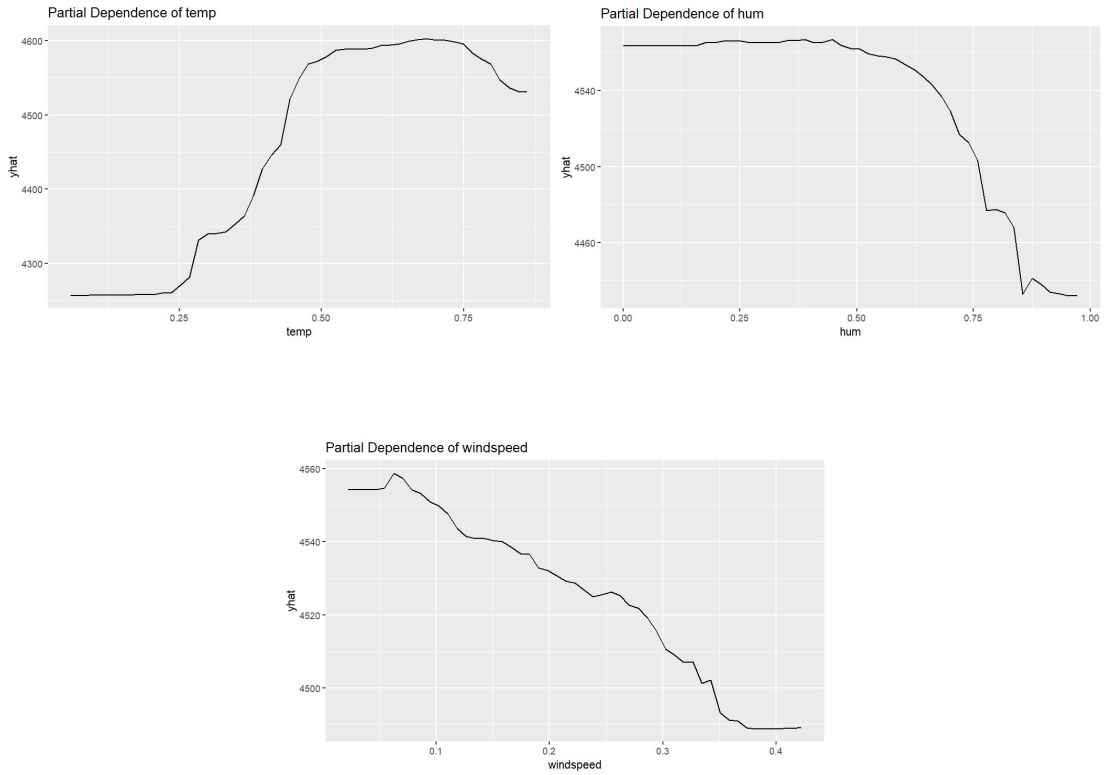
Ignacio Escriche Galbis

Almudena Peiró Blanco

Understanding the factors that influence predicted outcomes is crucial for effective data-driven decision-making. One useful tool for this is the Partial Dependence Plot (PDP), which shows the marginal effect of a feature on the predicted outcome.

In this analysis, we use PDPs to examine a regression model that predicts bike rentals. By fitting a random forest model, we visualize how environmental variables—temperature, humidity, wind speed, and days since 2011—affect the predicted bike counts.

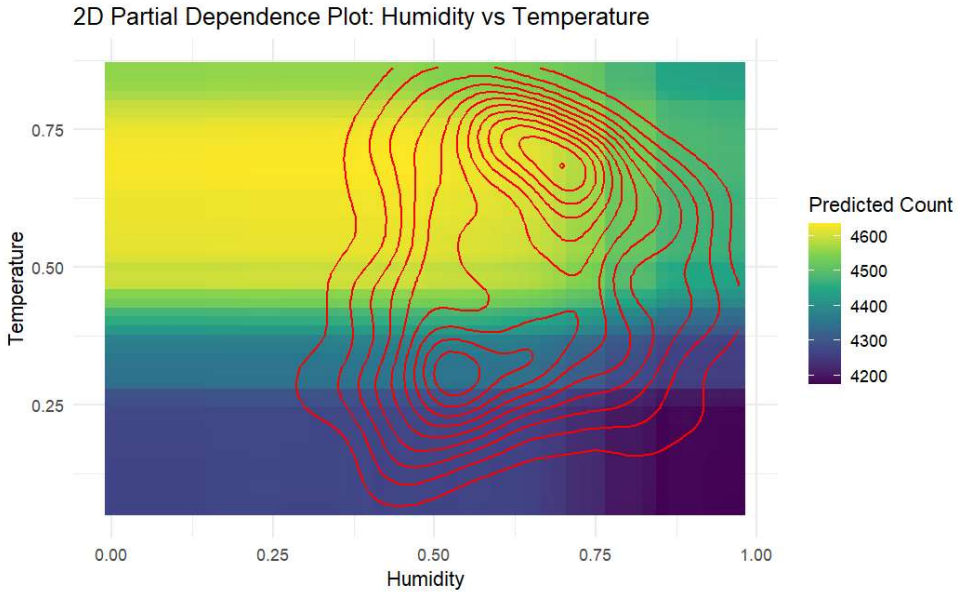
Following this, we will apply PDPs to understand the factors influencing house prices and explore Bidimensional Partial Dependency Plots to analyze interactions between two features and their combined effect on predictions.



The Partial Dependence Plots (PDPs) reveal several key insights about the influence of environmental factors on predicted bike counts. Temperature has a positive relationship with bike counts, with the number of rentals increasing as the temperature rises, peaking at a normalized value of around 0.6, after which extremely high temperatures lead to a slight decrease in rentals. Humidity shows an inverse relationship; bike counts remain stable at lower humidity levels but drop significantly as humidity exceeds 0.6, likely due to discomfort from high humidity. Wind speed also negatively impacts bike counts, with higher wind speeds leading to fewer rentals, as biking becomes more challenging.

These insights highlight how temperature, humidity, and wind speed affect biking activity, providing valuable information for managing bike rental services.

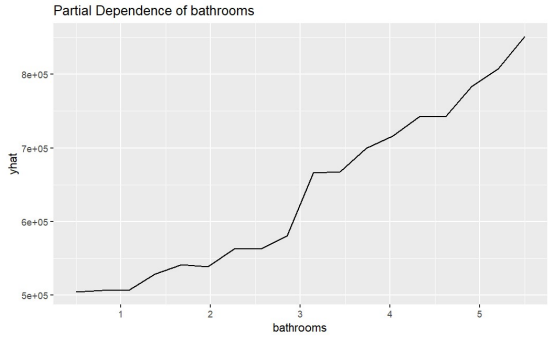
Using a 2D Partial Dependence Plot (PDP) instead of solely relying on 1D PDPs provides a more comprehensive understanding of how two features (for example humidity and temperature) interact to influence the predicted outcome (count of bikes). While 1D PDPs offer insights into the marginal effect of individual features, a 2D PDP allows us to visualize the joint impact of these two features, revealing complex relationships and interactions that may not be apparent when considering each feature in isolation. By plotting multiple dimensions simultaneously, we gain deeper insights into the nuances of the model's behavior and can identify optimal conditions or critical thresholds where the predicted outcome is most affected.

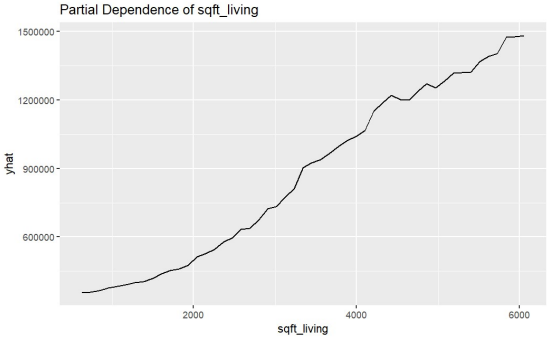


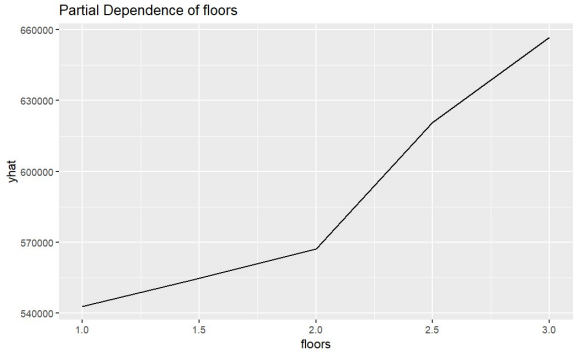
The 2D Partial Dependence Plot of humidity versus temperature for predicting bike counts shows that the optimal conditions for high bike rentals occur at moderate to high temperatures (around 0.5 to 0.75 normalized values) with low to moderate humidity levels (below 0.5). The plot reveals that as humidity increases, the predicted bike counts decrease significantly, even if the temperature is favorable. This indicates that humidity is a stronger deterrent to biking than temperature. The highest predicted bike counts are observed in areas with moderate temperatures and low humidity, while high humidity levels consistently result in lower bike counts. These insights suggest that favorable biking conditions are primarily driven by moderate temperatures and lower humidity, highlighting the importance of these factors in predicting bike rental demand.

PDPs have provided us valuable insights predicting bike rentals, similarly, they can be applied to explain the price of a house. The Partial Dependence Plots (PDPs) for the house price prediction model reveal several key insights.

id <chr>	date <S3: POSIXct>	price <dbl>	bedrooms <dbl>	bathrooms <dbl>	sqft_living <dbl>	sqft_lot <dbl>	floors <dbl>	waterfront <dbl>
7129300520	2014-10-13	221900	3	1.00	1180	5650	1	0
6414100192	2014-12-09	538000	3	2.25	2570	7242	2	0
5631500400	2015-02-25	180000	2	1.00	770	10000	1	0
2487200875	2014-12-09	604000	4	3.00	1960	5000	1	0
1954400510	2015-02-18	510000	3	2.00	1680	8080	1	0
7237550310	2014-05-12	1225000	4	4.50	5420	101930	1	0







Firstly, the number of bedrooms has an inverse relationship with the predicted house price, particularly beyond two bedrooms, where an increase in bedrooms leads to a decrease in price. This may indicate that excessively large houses with many bedrooms are less valued. In contrast, the number of bathrooms shows a positive correlation with house price; more bathrooms generally correspond to higher house prices, reflecting the added luxury and convenience. Similarly, the square footage of living space (sqft_living) exhibits a strong positive relationship with house price, with larger living spaces significantly boosting the predicted value. Lastly, the number of floors also positively influences house prices, especially moving from one to two floors, suggesting that multi-story homes are considered more desirable and valuable. These findings highlight the importance of certain features—bathrooms, living space, and the number of floors—in enhancing

property value, while an excessive number of bedrooms might not add proportional value.

In conclusion, Partial Dependence Plots are valuable tools for understanding how different features influence the predictions of regression models.

In both the context of bike rentals and house price predictions, these analyses provide critical insights that can guide decision-making and optimize strategies based on the most influential conditions and features.

For bike rental services, this means adjusting operations and marketing to leverage favorable environmental conditions. For real estate markets, it involves optimizing property attributes to align with buyer preferences.