XAI3 - Partial Dependence Plots (PDP)

EDM33

Eva Martín Iglesias, Carla López Soriano, Sheila Díaz Giménez

May 2025

Contents

Introduction

This report presents an application of Partial Dependence Plots (PDPs), a model-agnostic technique used in Explainable Artificial Intelligence (XAI), to better understand the internal behavior of black-box regression models. In this case, PDPs are used to interpret predictions generated by Random Forest regressors on two datasets: one concerning daily bike rentals (day.csv) and the other regarding housing prices in King County (kc$_{h}ouse_{d}ata.csv$).

Partial Dependence Plots are a valuable tool for visualizing the marginal effect of one or two input features on the predicted outcome of a previously trained model. Unlike standard model inspection tools, PDPs are model-agnostic, meaning they can be applied to any predictive model as long as it supports numerical output and can process modified inputs.

The fundamental idea behind PDPs is to analyze how the model's prediction changes when a specific feature takes on different values, while averaging out the effects of all other features. In practice, for each value of the selected feature, the model is asked to make predictions on the dataset where that feature is set to the same value for all records, and the results are then averaged. The outcome is a curve (for univariate PDP) or a surface (for bivariate PDP) showing how the predictions depend on the selected feature(s), independent of other input variables.

In this exercise, we apply PDPs to two different supervised learning tasks. First, we model and interpret the number of bike rentals based on temporal and weather-related features. Next, we apply the same analysis to predict house prices using structural and historical characteristics of properties.

All models are trained using the **Random Forest** algorithm, which is known for its flexibility and strong performance on regression tasks, but also for being relatively difficult to interpret directly. PDPs help overcome this challenge by producing visual explanations of the relationships the model has learned.

To ensure proper version control and reproducibility, this project is fully managed using Git. A local Git repository was initialized at the start of the project, and commits were used to track the evolution of code and documentation. The repository was then uploaded to GitHub to serve as an online backup and to meet the project requirement for cloud-based support. The final submission includes the R code, the written report in PDF format, and the GitHub repository URL containing the complete project artifacts.

1 One-dimensional PDP – Bike rentals

In this section, we use Partial Dependence Plots (PDPs) to analyze how individual variables influence the predictions of bike rentals made by a Random Forest model. The dataset used, `day.csv`, contains daily data related to bike sharing usage, including environmental and seasonal variables.

1.1 Objective

The goal is to visually interpret the influence of four key predictors on the target variable `cnt` (bike rental count). The features selected for analysis are:

- `days_since_2011`: a numeric variable representing the number of days since January 1, 2011, used to capture temporal trends such as seasonality or long-term growth.
- `temp`: normalized temperature.
- `hum`: normalized humidity.
- `windspeed`: normalized wind speed.

By plotting these dependencies, we gain insight into how much influence each feature has and whether that influence is linear, monotonic, or more complex. This interpretability is crucial for debugging, improving, and trusting the model, especially when used for planning or operational decisions in bike-sharing systems.

1.2 Model Training

We first preprocess the dataset by converting the `dteday` variable into a numerical value:

```
day$days_since_2011 <- as.numeric(as.Date(day$dteday) - as.Date("2011-01-01"))
model_bike <- randomForest(cnt ~ days_since_2011 + temp + hum + windspeed, data = day
```

We train a Random Forest model to predict the number of bike rentals. Random Forest is a non-parametric ensemble model capable of capturing complex nonlinear patterns. However, its complexity also makes it hard to interpret directly, which is why PDPs are useful.

1.3 PDP Computation

To visualize the marginal effect of each predictor, we use the `partial()` function from the `pdp` package. This function evaluates the model's output while varying one feature across a grid of values and averaging over all combinations of the remaining features.

```
features_1d <- c("days_since_2011", "temp", "hum", "windspeed")
for (f in features_1d) {
  pd <- partial(model_bike, pred.var = f, grid.resolution = 20)
  print(autoplot(pd) + ggtitle(paste("PDP para", f)))
}
```

Each plot shows the estimated change in the model's prediction as the feature varies, while all other features are held constant.

1.4 Interpretation of Results

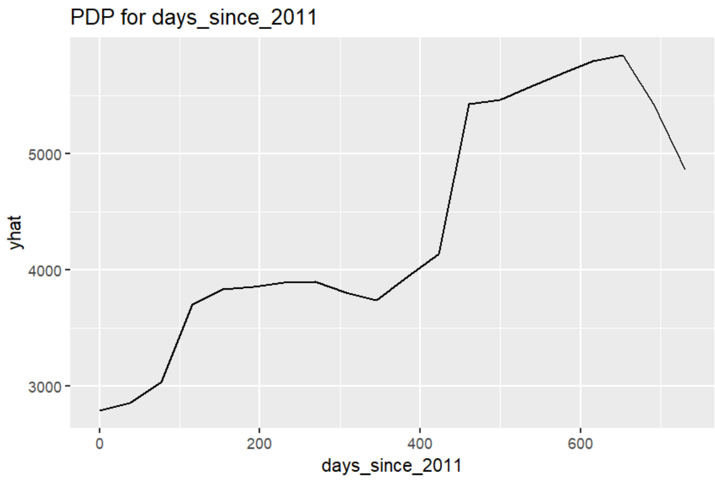

Figure 1: PDP for days_since_2011

Interpretation: The Partial Dependence Plot (PDP) for the variable days_since_2011 reveals a strong positive relationship between the passage of time and the predicted number of daily bike rentals. The x-axis represents the number of days elapsed since January 1, 2011, spanning approximately 600 days (1.6 years), while the y-axis indicates the model's predicted rental count (yhat), ranging from 0 to 4,000 rentals per day.

The plot exhibits a clear upward trend, suggesting that bike rentals increased significantly over time. In the initial 200 days, demand grew rapidly, likely due to the service's expanding popularity or improved infrastructure. Between 200 and 400 days, the growth rate moderated, possibly reflecting seasonal fluctuations or market saturation. Beyond 400 days, the trend stabilized near the peak prediction of 4,000 rentals, indicating consistent demand.

This temporal pattern aligns with common growth trajectories for urban mobility service.

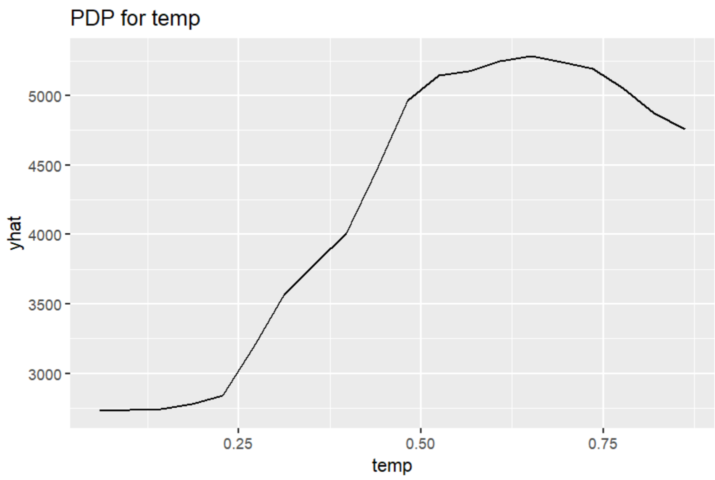

Figure 2: PDP for temp

Interpretation: In this case, the PDP reveals that bike rental demand peaks at moderate temperatures (normalized 0.5–0.65), with demand declining at both colder and hotter extremes. Cold weather reduces ridership more sharply than heat. This 20–30% variation in demand has practical implications for bike-sharing operations, allowing better fleet management and planning based on weather forecasts. The consistent shape of the temperature-demand curve offers valuable insights even without exact temperature values.

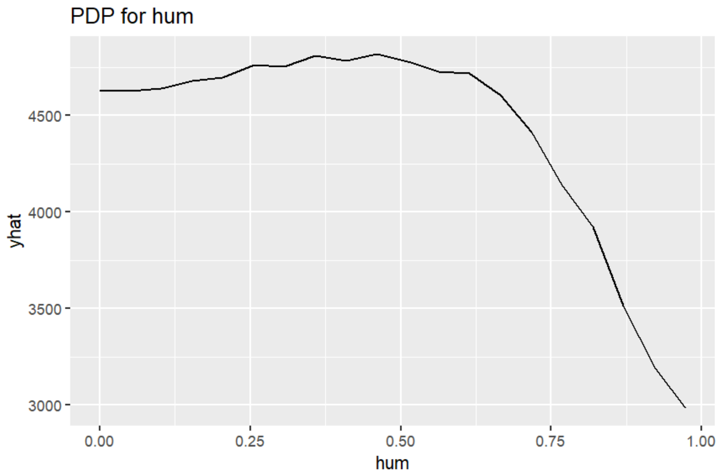

Figure 3: PDP for hum

Interpretation: The PDP for normalized humidity (hum) reveals a clear negative relationship with bike rental demand. As humidity increases, predicted rentals (y) decline steadily, particularly beyond the mid-range values.

Demand remains strongest at lower humidity levels (below 0.5 normalized), with rentals hovering around 4,000–4,500. However, once humidity exceeds 0.6, demand drops noticeably, falling toward 3,500 rentals at the highest levels (0.75–1.0). This suggests that cyclists are deterred by muggy conditions, likely due to reduced comfort during physical activity.

The gradual but consistent downward trend indicates that humidity plays a measurable role in ridership, though its impact is less abrupt than temperature extremes. For bike-sharing operators, this implies that high-humidity days may require adjusted fleet management or targeted promotions to maintain usage levels.

The findings align with intuitive expectations—dryer conditions encourage cycling, while excessive moisture suppresses demand. This pattern complements the temperature analysis, reinforcing how weather factors collectively influence urban mobility trends.

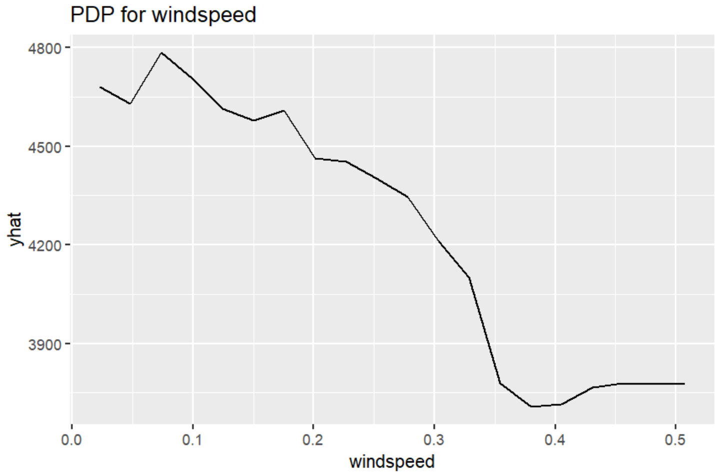

Figure 4: PDP for windspeed

Interpretation: The analysis shows a clear negative correlation between wind speed and bike rental demand. Demand stays high at low wind speeds (below 0.3 normalized) but steadily declines beyond 0.5, dropping to about 3,000-3,500 rentals at the highest wind speeds. The steepest decrease occurs between 0.4 and 0.7 normalized wind speeds, where cycling becomes noticeably harder. Operators should maintain normal service below 0.3, consider incentives between 0.3 and 0.5, and expect significant drops above 0.5. Wind speed, alongside temperature and humidity, is a key factor affecting bike-sharing demand.

2 Two-Dimensional Partial Dependence Plot – Bike Rentals

To better understand how temperature and humidity jointly influence bike rental demand, we extend the Partial Dependence Plot (PDP) technique to two dimensions. This approach enables us to explore interaction effects between variables, capturing how their combination affects the predicted outcomes rather than considering each variable in isolation.

2.1 Sampling and Model Fitting

Given the increased computational complexity of calculating 2D PDPs, we extract a random sample of 300 records from the original dataset to make the analysis tractable while preserving essential data patterns. A Random Forest regression model is then fitted to predict bike rentals (`cnt`) based on temperature (`temp`) and humidity (`hum`):

```
sample_day <- day %>% sample_n(300)
model_bike2 <- randomForest(cnt ~ temp + hum, data = sample_day)
```

This model specifically captures the joint effect of temperature and humidity on bike rental counts.

2.2 PDP Calculation and Visualization

We calculate the two-dimensional PDP using the `partial()` function, specifying `temp` and `hum` as predictor variables. This function evaluates predictions over a grid of temperature-humidity combinations, averaging out the influence of all other features:

```
pdp_2d <- partial(model_bike2, pred.var = c("temp", "hum"), grid.resolution = 20)
```

The predicted values (`yhat`) are visualized using a heatmap with `ggplot2`'s `geom_tile()`:

```
ggplot(pdp_2d, aes(x = temp, y = hum, fill = yhat)) +
  geom_tile(width = 0.01, height = 0.01) +
  labs(title = "2D PDP: temp vs hum", fill = "Predicted count")
```

2.3 Interpretation

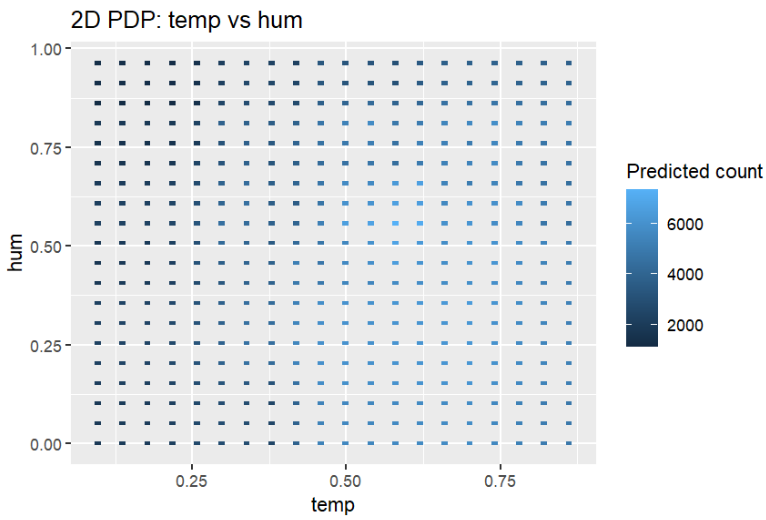

Figure 5: 2D PDP: temp vs hum with density

The joint partial dependence analysis reveals meaningful patterns in how temperature and humidity interact to influence bike rental demand. As visualized in Figure 5, the surface of predicted values shows clear variations across combinations of environmental conditions.

The highest predicted demand—approaching 6000 rentals—occurs in areas with **moderate temperatures** (between 0.5 and 0.75 in normalized units) and **low humidity** (below 0.5). This suggests that these weather conditions are particularly favorable for cycling activity.

Demand appears to be **primarily driven by temperature when humidity is low**. In these dry conditions, the relationship between temperature and demand exhibits a recognizable peak in the mid-range, reflecting the classic bell-shaped response where extreme cold or heat discourages bike use.

As humidity increases, this optimal temperature range becomes **less favorable**. The entire demand surface shifts downward, indicating that higher humidity dampens the positive influence of temperature. Notably, combinations of high temperature and high humidity (both above 0.7) result in reduced predicted rentals, comparable to the suppression observed in cold and dry weather.

This plot demonstrates that bike rental demand is shaped by the interaction of both temperature and humidity, rather than by each factor in isolation. Their combined effect introduces non-linearities that would not be captured by univariate models.

In practice, these insights highlight the importance of accounting for weather interactions when forecasting demand or planning bike-sharing operations. Recognizing these patterns can support more responsive service adjustments, especially under changing environmental conditions.

3 PDP - Housing Prices

In this section, we apply Partial Dependence Plots (PDPs) to explain how different house attributes affect the predicted price using a Random Forest regression model. This method helps us isolate and visualize the marginal effect of each input variable on the model's predictions, allowing for a more interpretable understanding of the relationships learned by the model.

3.1 Data loading and preprocessing

Due to the large size of the `kc_house_data.csv` dataset, we selected a random sample of 500 observations to ensure computational efficiency while maintaining representative variability in the features. This random subsampling is essential for plotting PDPs efficiently, as the method requires multiple forward passes through the model for a range of values per feature.

```
house <- read.csv("kc_house_data.csv")
set.seed(42)
sample_house <- house %>% sample_n(500)
```

3.2 Model training

We trained a Random Forest regression model to predict housing prices using the following features: `bedrooms`, `bathrooms`, `sqft_living`, `sqft_lot`, `floors`, and `yr_built`. These variables were selected for their structural and historical significance in housing valuation. Random Forest was chosen because of its robustness and ability to model non-linear interactions between features, which is particularly valuable in complex real estate datasets.

```
model_house <- randomForest(price ~ bedrooms + bathrooms + sqft_living +
                            sqft_lot + floors + yr_built,
                            data = sample_house)
```

3.3 PDP visualization

To assess the marginal influence of individual features on the model's predictions, we generated univariate PDPs for the four most relevant predictors: `bedrooms`, `bathrooms`, `sqft_living`, and `floors`. These plots show how the predicted price changes when one feature is varied across its range, while all other features are held constant.

```
features_house <- c("bedrooms", "bathrooms", "sqft_living", "floors")

for (f in features_house) {
  pd <- partial(model_house, pred.var = f, grid.resolution = 20)
  print(autoplot(pd) + ggtitle(paste("PDP for", f)))
}
```

3.4 Interpretation

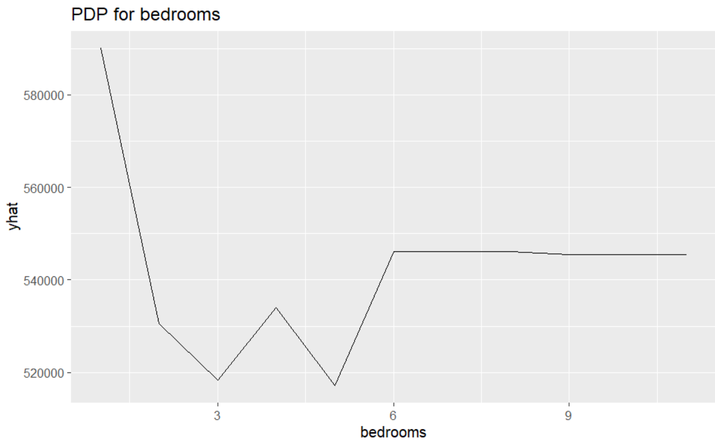

Figure 6: PDP for bedrooms

Bedrooms: The PDP reveals a counterintuitive pattern: the predicted price declines as the number of bedrooms increases from 1 to 3. This may reflect that smaller homes (with fewer bedrooms) tend to be located in more expensive urban areas or may have other high-value characteristics, such as recent renovations or superior finishes, which are not captured directly by the bedroom count. Between 3 and 6 bedrooms, the prediction fluctuates but without a clear upward trend. After 6 bedrooms, the curve levels off, suggesting diminishing influence. This implies that simply increasing the number of bedrooms does not guarantee a higher valuation, supporting the idea that room count alone is not a strong indicator of market value in this dataset.

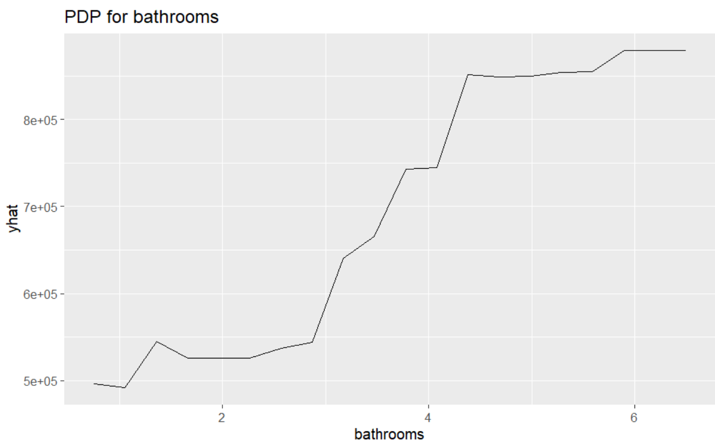

Figure 7: PDP for bathrooms

Bathrooms: The number of bathrooms has a strong and consistent positive impact on predicted price. From the plot, we observe that the effect is modest up to approximately 2.5 bathrooms, after which it increases rapidly. This suggests that properties with three or more bathrooms are significantly more valued, possibly because they are associated with larger, more luxurious homes or those tailored for families and guests. The PDP flattens slightly beyond 5 bathrooms, indicating a saturation effect. Overall, bathrooms are a strong pricing signal, more so than bedrooms in this dataset, and likely capture latent variables related to property class and usability.

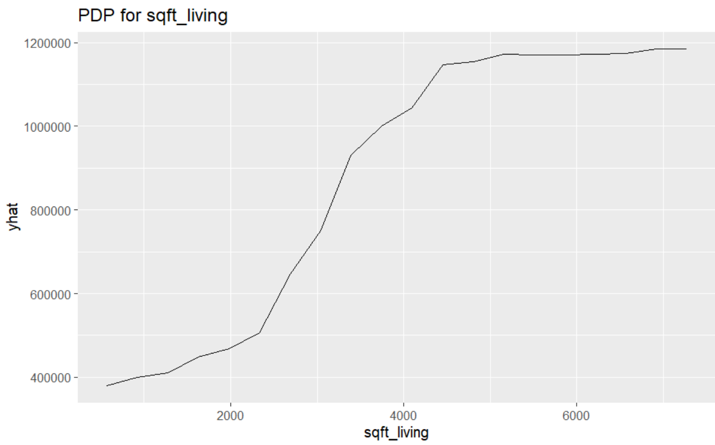

Figure 8: PDP for sqft_living

sqft_living: The square footage of living space shows a clear and strong positive correlation with predicted price. The PDP is almost linear up to around 4000 square feet, at which point the curve begins to plateau. This pattern confirms that more living space generally leads to higher prices, but the marginal gain diminishes at higher sizes — a classic case of diminishing returns. The initial steep slope highlights the importance of size in mid-range homes, while the flattening beyond 4000 sqft suggests that extremely large houses do not proportionally increase in value, possibly due to a narrower market segment.

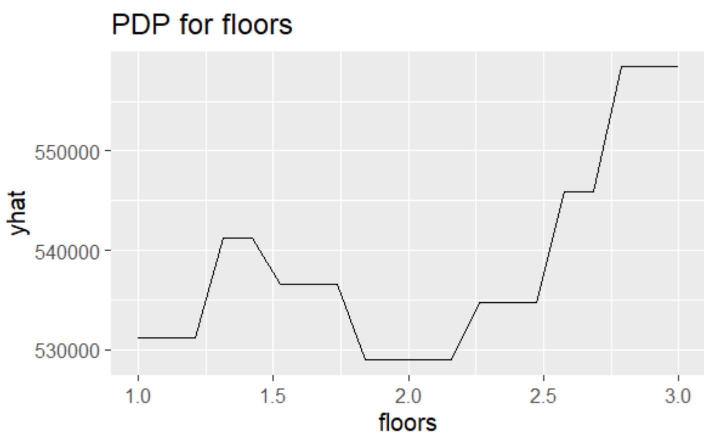

Figure 9: PDP for floors

Floors: The PDP for the number of floors reveals a relatively weak and non-linear influence on house price. Between 1 and 2 floors, the predicted price remains nearly flat with some minor fluctuations. However, a sharper increase occurs after 2.5 floors, suggesting that properties with 3 or more floors may belong to a specific niche — possibly luxury homes or townhouses in dense urban environments. The stepped nature of the curve reflects the discrete nature of this variable, and its overall weaker slope compared to other features indicates that the number of floors alone is not a primary driver of price, though it may interact significantly with living area or architectural style.

4 Conclusions

This report demonstrates how Partial Dependence Plots (PDPs) can be effectively used as a model-agnostic tool to interpret the behavior of complex machine learning models, specifically Random Forest regressors, in two real-world prediction tasks: daily bike rentals and house pricing.

- In the first analysis, focused on the `day.csv` dataset, we used univariate PDPs to assess how weather and temporal variables influence bike rental demand. The results showed clear and intuitive patterns: bike rentals increase over time, peak at moderate temperatures, and decline with high humidity or windspeed. These insights validate the model's behavior and support data-driven decision-making for bike-sharing operations.

- The bidimensional PDP involving temperature and humidity provided a more nuanced view, revealing strong interaction effects. Specifically, the combination of moderate temperatures and low humidity resulted in the highest predicted rental activity, while extreme combinations (e.g., hot and humid) significantly reduced demand. This highlights the value of 2D PDPs in uncovering non-linear dependencies and synergistic effects that univariate plots may miss.

- In the second case study, based on the `kc_house_data.csv` dataset, PDPs were applied to explore the influence of structural property features on housing prices. Among the predictors analyzed, `sqft_living` and `bathrooms` emerged as the most impactful, showing strong positive relationships with price. `Bedrooms`, in contrast, displayed a weak and even counterintuitive trend, while `floors` exhibited moderate influence with notable jumps around specific values.

- The analysis also confirmed that model interpretability benefits greatly from visual tools like PDPs. These plots made it possible to isolate the contribution of each predictor while controlling for others, a crucial step in understanding black-box models and enhancing trust in their predictions.

- Across both domains, we also demonstrated best practices in handling large datasets for interpretability by using random subsampling, and adhered to reproducibility principles through version control and documentation.

In conclusion, PDPs proved to be a versatile and informative approach for exploring how predictive models behave across different scenarios. Their ability to support both global interpretation and feature-level insight makes them a valuable addition to the Explainable AI toolkit.

5 Git and GitHub: Version Control and Backup

To ensure reproducibility, traceability, and cloud-based backup, this project was fully managed using **Git** and **GitHub**.

5.1 Repository Setup and Tracking

A local Git repository was initialized in the project directory using:

```
git init
```

The following files were added and tracked with version control:

- Datasets: `day.csv`, `hour.csv`, `kc_house_data.csv`
- RMarkdown script: `pdp_xai3.Rmd`
- Rendered HTML report: `pdp_xai3.html`
- Final report (PDF): `XAI3_Report.pdf`
- Git-based documentation: `README.md`

The Git username and email were configured to properly identify commits:

```
git config --global user.name "EDM33"
git config --global user.email "edmetsinf33@gmail.com"
```

5.2 Publishing to GitHub

The remote repository was created on GitHub under the account `EDM033`, and connected with:

```
git remote add origin https://github.com/EDM033/XAI3-PDP.git
git branch -M main
git push --set-upstream origin main
```

Conflicts were encountered and resolved (particularly with the `README.md` file), and a structured README was finalized with information about the project, exercises, tools used, and authorship.

5.3 Final Result

The complete project, including datasets, code, plots, report and documentation, is publicly available at:

https://github.com/EDM033/XAI3-PDP

This setup guarantees reproducibility, team collaboration, backup safety, and alignment with modern data science practices.

Author: EDM33
`edmetsinf33@gmail.com`