

Informe de Resultados: Partial Dependence Plots (PDP)

XAI - Métodos Agnósticos al Modelo

Ejercicio 5 - Visualización y Análisis con PDP

Repositorio GitHub: <https://github.com/raseval/ejercicio-pdp>

1. PDP unidimensional – Predicción de Alquiler de Bicicletas

Se ha entrenado un modelo Random Forest para predecir el número de bicicletas alquiladas (`cnt`) y se han generado gráficos PDP para analizar el efecto marginal de las siguientes variables:

a) `days_since_2011`

- El gráfico muestra un crecimiento progresivo en la predicción de alquileres conforme avanza el tiempo desde 2011.
- Esto sugiere una tendencia creciente en el uso de bicicletas a lo largo del periodo, posiblemente debido a factores como mayor conciencia ecológica, mejora de infraestructuras o cambios en patrones de movilidad urbana.

b) `temp` (temperatura normalizada)

- Se observa una fuerte relación positiva hasta valores cercanos a 0.75, donde el número de alquileres alcanza su pico.
- A partir de ese punto, el aumento adicional de temperatura genera una leve caída o estabilización en la predicción.
- Esto puede interpretarse como que temperaturas moderadas favorecen el uso de la bicicleta, mientras que temperaturas extremas no lo hacen.

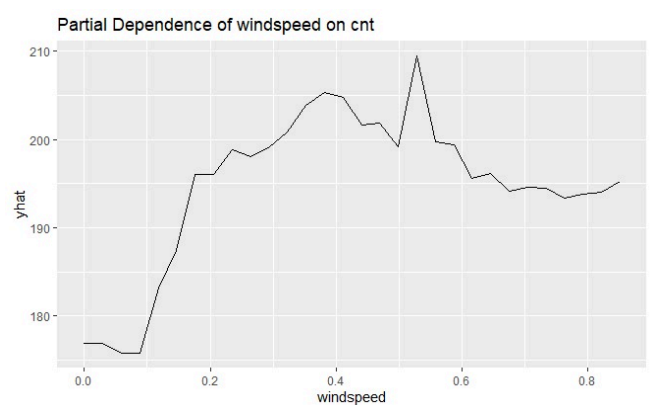
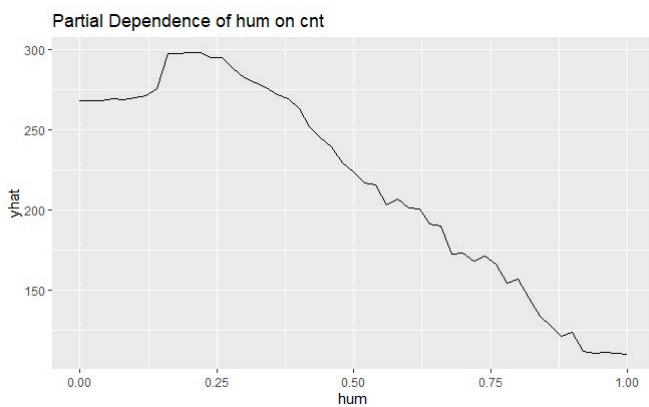
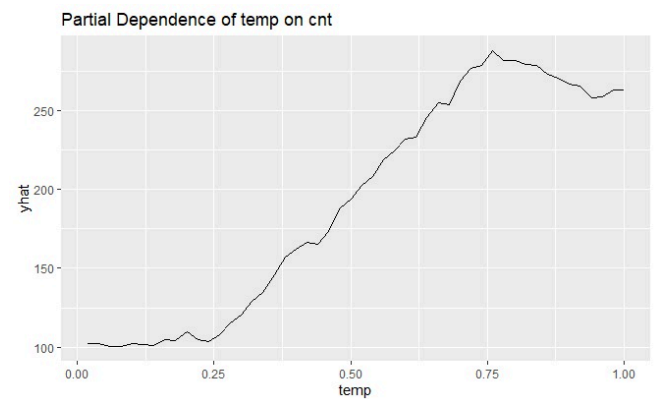
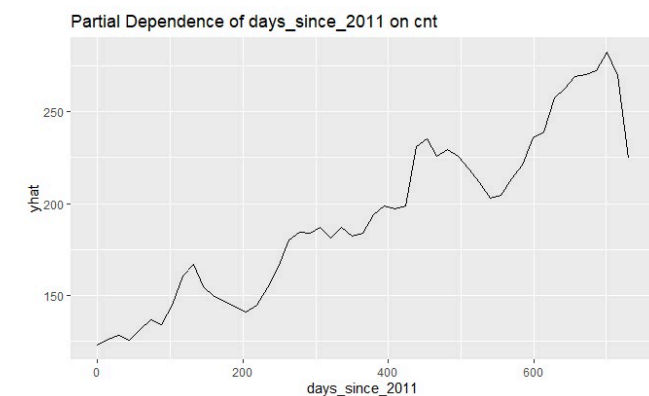
c) `hum` (humedad)

- Existe una clara relación inversa: a mayor humedad, menor es el número de alquileres predichos.

- Este resultado es esperable, ya que altos niveles de humedad tienden a coincidir con climas incómodos o lluviosos, poco favorables para el uso de bicicletas.

d) **windspeed** (velocidad del viento)

- El efecto es más tenue, pero se identifica una leve disminución en el número de alquileres para velocidades de viento altas.
- Un pico inusual aparece alrededor de 0.55, probablemente debido a ruido o a una baja densidad de datos en ese rango.

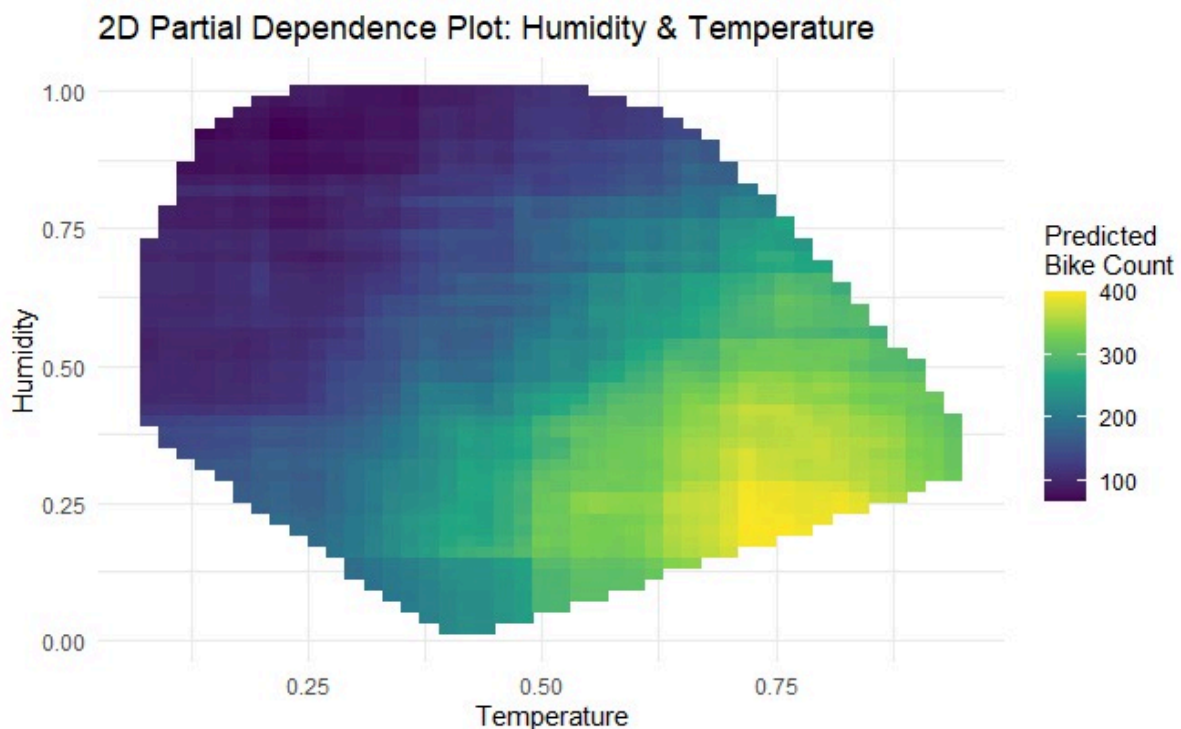


2. PDP Bidimensional – Humedad y Temperatura

Se construyó un gráfico de dependencia parcial bidimensional utilizando `humidity` y `temperature`.

Observaciones:

- Las zonas de mayor predicción de alquileres se encuentran en condiciones de temperatura media-alta y baja humedad.
- Los niveles más bajos de predicción aparecen en ambientes muy húmedos (arriba a la izquierda del gráfico), sin importar la temperatura.
- Esto refuerza la conclusión de que la humedad es una variable limitante importante para el alquiler de bicicletas.



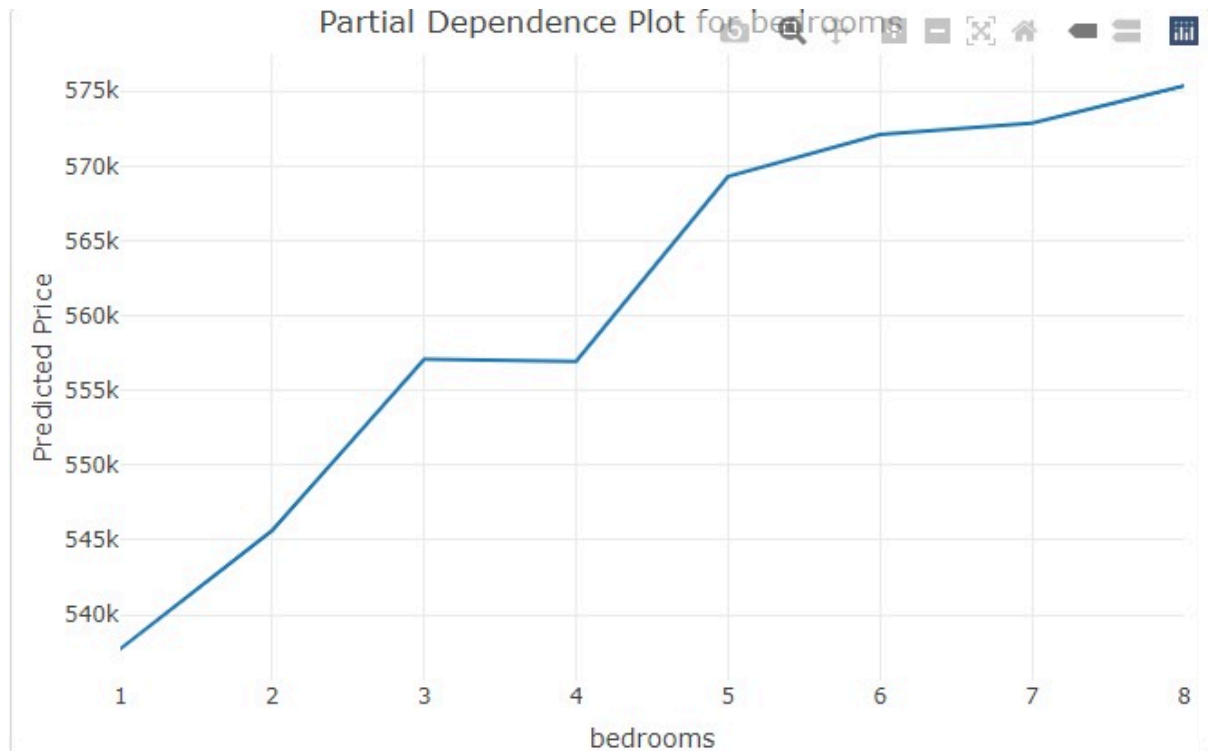
3. PDP – Predicción del Precio de Viviendas

Usando el dataset `kc_house_data.csv`, se entrenó un modelo Random Forest para predecir el precio (`price`) en función de varias variables estructurales.

Se analizaron las siguientes variables mediante PDP:

a) bedrooms

- Existe una correlación positiva moderada: el precio predicho aumenta con el número de habitaciones.
- Sin embargo, el incremento se reduce a partir de 5 habitaciones, indicando una posible saturación del efecto de este atributo.



Conclusión General

Los gráficos de dependencia parcial han permitido visualizar de forma clara y explicativa el impacto de distintas variables en modelos de regresión entrenados con Random Forest. En ambos casos analizados (uso de bicicletas y precios de vivienda), las variables ambientales o estructurales tienen efectos previsibles y coherentes con la intuición y el dominio del problema.