

# Explainable AI Practical Report III

*María Verdú Gómez y Hugo Sánchez Navalón*

## Índice

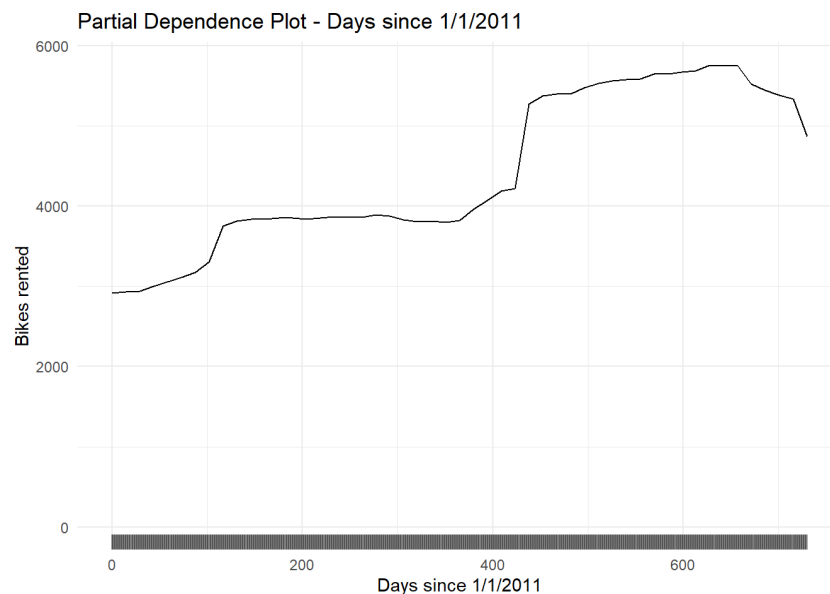
Bikes rental.....	2
Partial Dependence Plot unidimensional.....	2
Días desde el 1 de enero de 2011.....	2
Temperatura.....	3
Humedad.....	3
Velocidad del viento.....	4
Partial Dependence Plot bidimensional.....	5
House pricing.....	5
Dormitorios.....	6
Baños.....	6
Superficie habitable.....	7
Plantas.....	7

# Bikes rental

## Partial Dependence Plot unidimensional

Tras elaborar un modelo Random Forest para predecir la cantidad de bicicletas que se espera que sean demandadas en un servicio de alquiler de bicicletas, llevamos a cabo diversos gráficos de dependencia parcial para explicar cómo afecta la variación de distintas variables en la cantidad final de bicicletas.

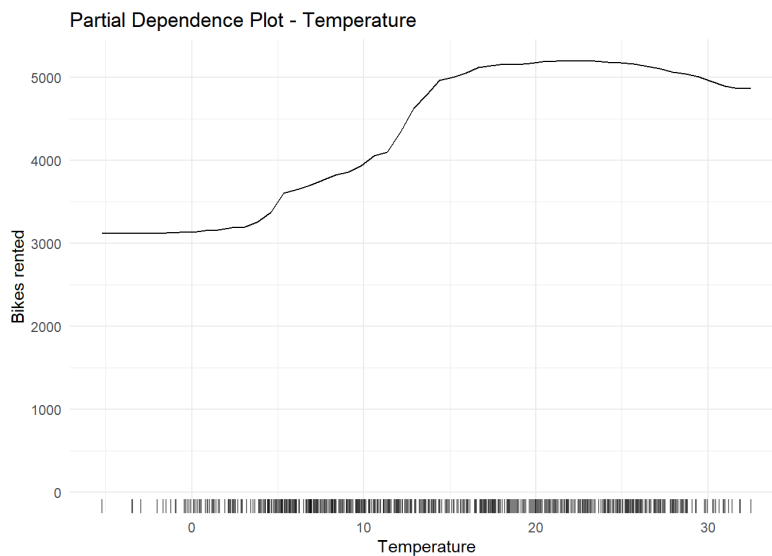
### Días desde el 1 de enero de 2011



En este gráfico puede interpretarse como la tendencia registrada en el servicio de bicicletas a demanda a igualdad de resto de variables observadas, puesto que la distribución ofrece una observación por día desde el 1 de enero de 2011. Observamos un comienzo en torno a las 3000 bicicletas ese día, que aumentan principalmente en dos grandes saltos; uno de ellos tiene lugar pasados algo más de 100 días desde el 1 de enero de 2011 y sitúa la cantidad de bicicletas esperada en algo menos de 4000 y el otro tiene lugar pasados los 400 días. En este último, la cantidad de bicicletas esperadas a igualdad del resto de parámetros supera las 5000 y, desde ese momento, crece paulatinamente hasta rozar las 6000 poco después de pasados los 600 días desde enero de 2011, registrándose después una leve caída.

Podemos concluir que el efecto recogido de la variable que describe los días pasados desde el 1 de enero de 2011 deja entrever que la tendencia general del uso del servicio de bicicletas ha sido creciente pero no de forma monótona sino probablemente debido a dos fenómenos puntuales concretos.

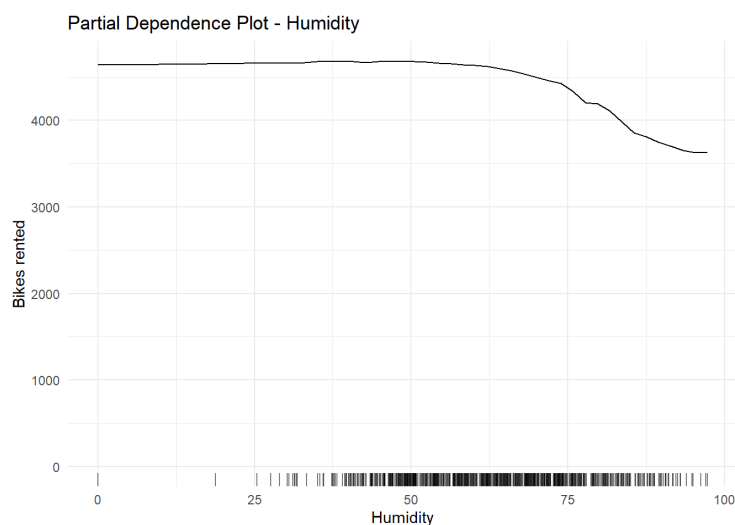
## Temperatura



Podemos observar que para valores menores a 0°C o superiores a 30°C disponemos de muy pocas observaciones, por lo que estaríamos obteniendo información sesgada si interpretáramos dichas regiones del gráfico. Si nos centramos en el intervalo de 5 a 28°C, donde se concentran la mayoría de estas, vemos que tiene lugar un incremento sostenido de la demanda de bicicletas por cada grado centígrado que sube la temperatura hasta alcanzar unos 16°C. A partir de entonces, no parece haber una variación sensible en el número de bicicletas demandadas esperadas suponiendo que el resto de parámetros se mantienen constantes.

La principal conclusión es que, en días de notorio frío, cabe esperar una reducción de la cantidad de bicicletas demandadas acorde a la intensidad del frío que se dé, suponiendo que no existen otros factores que perturben las asunciones respecto al resto de parámetros, que se espera que se mantengan constantes.

## Humedad

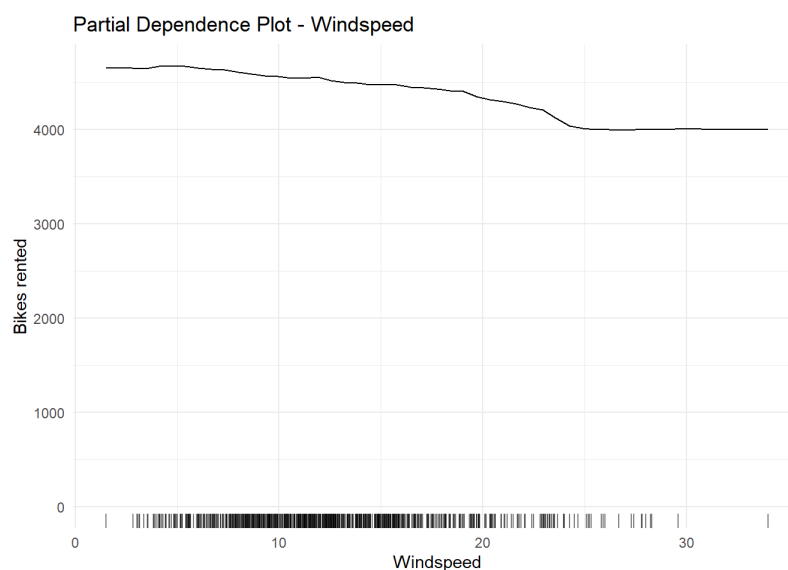


Para el caso de la humedad, observamos de igual manera que la distribución de las observaciones da lugar a regiones del gráfico cuya interpretación sería poco fiable. La inmensa mayoría de las

observaciones se concentra entre el 40 y el 80% de humedad. Para tal tramo, la cantidad de bicicletas demandadas parece mantenerse estable hasta el 70% de humedad, valor desde el que comienza un paulatino descenso de las bicicletas demandadas según aumenta la humedad en puntos porcentuales.

Podemos extraer como conclusión que, para niveles de humedad especialmente altos, se espera un descenso del número de bicicletas demandadas. Sin embargo, hemos de tener cuidado al afirmarlo, puesto que asumimos que el resto de variables se mantienen constantes y ese escenario, en el caso de la humedad, es poco probable. Los valores altos de humedad suelen estar relacionados con lluvias u otros fenómenos meteorológicos que pueden ser los que realmente estén detrás de este descenso de la cantidad demandada de bicicletas, por lo que cabe “coger con pinzas” el hecho de que la humedad, por sí sola, explique la demanda real.

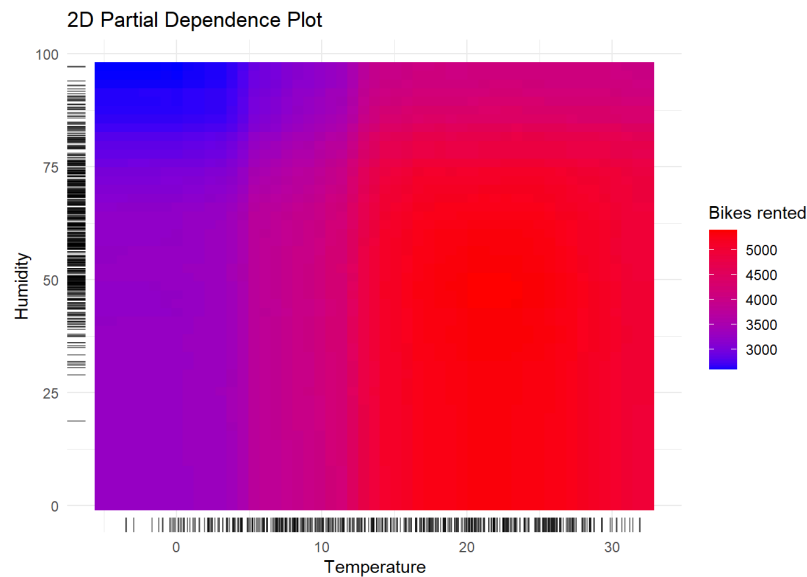
## Velocidad del viento



En el caso de la velocidad del viento, vemos que la mayor parte de valores disponibles oscilan entre los 3 y los 23 kilómetros por hora, aproximadamente. A lo largo de este intervalo, se registra una ligera tendencia decreciente en la cantidad demandada esperada, si bien esta no es muy grande.

Podemos concluir que, a igualdad del resto de variables observadas, los días en que el viento sea más veloz cabe esperar una oscilación a la baja del número de bicicletas demandadas esperado promedio.

## Partial Dependence Plot bidimensional



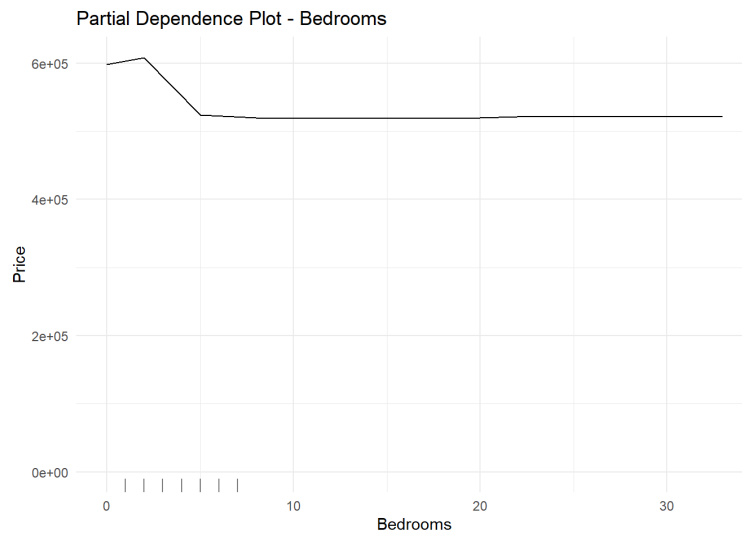
En esta visualización podemos observar el efecto conjunto de la temperatura y la humedad sobre el número de bicicletas alquiladas. Cabe decir que los valores de humedad están mayormente acotados entre el 40 y el 80%, lo que hace que la región inferior del gráfico no pueda interpretarse con gran precisión. De igual forma, el registro de los valores de temperatura hace que los laterales izquierdo y derecho carezcan de gran información hacia el conjunto de la interpretación.

Centrándonos en la parte que entendemos más fidedigna, que es la que excluye todas las áreas que acabamos de describir, vemos una relación inversa entre humedad y temperatura. Se espera que los días de alta humedad y baja temperatura el número de bicicletas sea menor, y aumente conforme la humedad decrezca y la temperatura aumente mientras permanezcan constantes el resto de variables observadas.

## House pricing

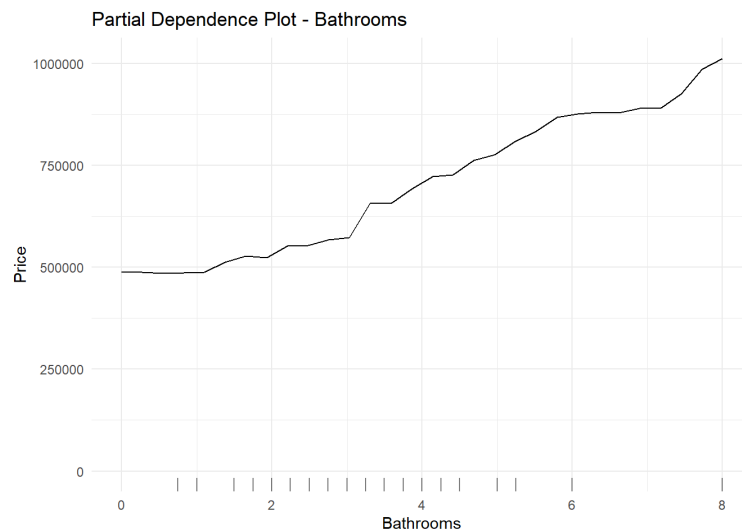
Elaboramos un modelo de regresión del precio de un conjunto de viviendas a partir de su número de baños, su número de dormitorios, su superficie habitable, su superficie total, su número de plantas y su año de construcción, empleando para ello un Random Forest. Posteriormente, tomamos una muestra de mil viviendas y aplicamos sobre ellas varios gráficos de dependencia parcial unidimensionales para explicar el efecto de diversas variables en el precio de una vivienda.

## Dormitorios



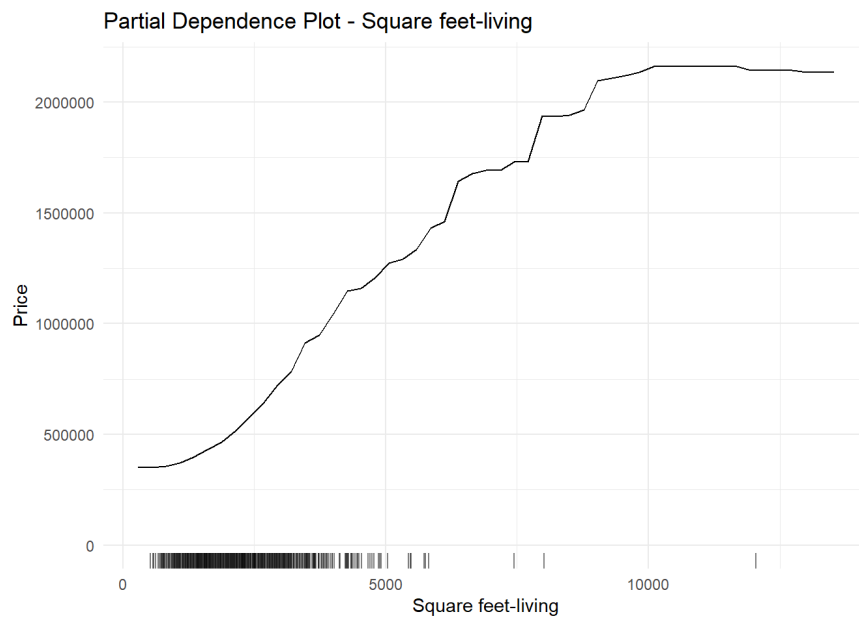
Observamos que el número de dormitorios produce una oscilación notable para valores menores a 5, que es donde tenemos un número de observaciones suficientes como para llegar a conclusiones plausibles. Se da la situación de que las viviendas con uno o dos dormitorios parecen ser más caras que aquellas que tienen más. Sin embargo, esto puede deberse a una escasez de observaciones con más de dos o tres dormitorios o a una eventual colinealidad con otros parámetros incluidos en el modelo de regresión, por lo que no hay gran evidencia que pueda extraerse de este gráfico.

## Baños



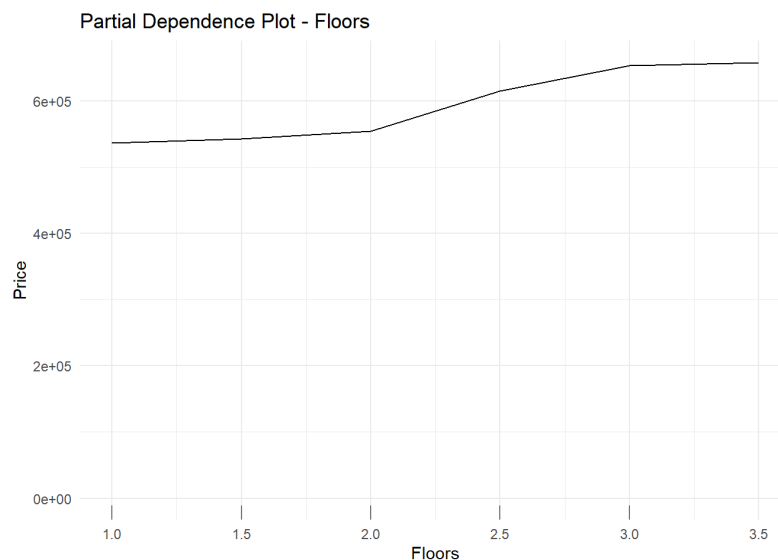
En el caso de los baños, observamos que disponemos de un número de observaciones suficiente como para poder interpretar el gráfico con un mínimo de fiabilidad para la región de 1 a 5. Observamos un crecimiento monótono del precio esperado de la vivienda a igualdad del resto de variable observadas conforme se añaden baños, por lo que podemos concluir que la instalación de cada baño adicional en una vivienda promedio encarece su precio en algunas decenas de miles de dólares.

## Superficie habitable



Para la región de la que más observaciones disponemos, acotada entre unos 500 y unos 4000 pies cuadrados, observamos un crecimiento que parece ser superlineal. Es decir, el coste de una vivienda promedio no aumenta de forma monótona al añadirle un pie cuadrado más de superficie habitable, sino que, conforme más superficie se construye, más se encarece la adición de más superficie habitable.

## Plantas



Observamos que disponemos de un rango poco amplio de valores observados, cuestión esperable si consideramos que resultan generalmente impracticables viviendas unifamiliares con más de tres plantas. Vemos que el precio de una vivienda de entre una y dos plantas no difiere demasiado, ampliándose la diferencia con la construcción de una tercera. En cualquier caso, la relación entre el

número de plantas y el precio es creciente, como cabría esperar de forma lógica para casas con características promedio.