



Universidad de San Andrés

BIG DATA

Condicionantes del precio de alquiler de departamentos pre y post Covid-19

Estudio de caso para la región del AMBA, Argentina

Mariana Bernad, Juan Andrés Cabral & Guido Sánchez

4 de enero de 2021

1. Introducción

Los determinantes de los precios de venta y alquiler de viviendas han sido ampliamente estudiados con el propósito de comprender y predecir la demanda del mercado inmobiliario. Utilizando modelos hedónicos, espaciales y métodos no paramétricos, distintos investigadores estimaron la importancia relativa de las distintas características de las viviendas y de las zonas en las que estas se encuentran. Sin embargo, las predicciones de los precios basadas en las características estudiadas son funcionales en la medida que se mantenga el contexto bajo las cuales fueron estudiadas.

La llegada del Covid-19 irrumpió este contexto, impactando la dinámica económica en todo el mundo. Entre las medidas de prevención más utilizadas se encuentran el aislamiento social, la cuarentena y el trabajo remoto. Las tres medidas implican un aumento, de distinta intensidad, en la proporción del día que se pasa dentro del hogar. Ante este cambio, nos preguntamos si las características que nos permiten predecir el precio de los inmuebles en alquiler varía pre y post Covid-19. En particular nos concentramos en las publicaciones inmobiliarias en el sitio Mercado Libre para la región del Área Metropolitana de Buenos Aires (AMBA), Argentina, entre 2019 y 2020.

El 20 de marzo del 2020 se declaró en la Argentina la cuarentena obligatoria por un período tentativo de 2 semanas. Este período terminó extendiéndose a 5 semanas para todo el país, seguido de una segmentación geográfica con restricciones que siguen hasta la fecha de hoy. Para la región del AMBA la cuarentena obligatoria se extendió hasta el 8 de noviembre, con una apertura de negocios gradual, pasando luego a una etapa de distanciamiento social. Según un informe del [INDEC](#) sobre el Mercado de Trabajo, la proporción de ocupados que trabaja desde su vivienda pasó del 6,5 % en el primer trimestre del 2020 al 22,0 % en el segundo trimestre del mismo año.

Ante este cambio de contexto, esperaríamos encontrar un cambio en las preferencias de ciertas características de los inmuebles demandados. En primer lugar, esperamos que aumente la demanda y el precio por metro cuadrado descubierto, es decir espacios al aire libre dentro de la vivienda, ya sea parque, patio o balcón. En segundo lugar, restringiéndonos a CABA, esperamos que aumente la demanda de alquileres lejos del Obelisco, proxy del centro de negocios de la ciudad, puesto que vivir cerca del trabajo deja de ser un beneficio. Por último, estudiaríamos los cambios sobre el precio de inmuebles que cuentan con pileta, gimnasio y servicios similares, por un lado el precio podría bajar en el corto plazo debido a la prohibición de su uso, pero por el otro, el precio podría aumentar en el largo plazo debido a la sustitución de establecimientos concurridos. Para responder estas preguntas llevaríamos a cabo un modelo CART para captar no linealidades en los predictores de los determinantes del precio y un modelo LASSO para obtener una lista reducida de predictores relevantes.

Analizando la base de datos del mercado inmobiliario proveída por Mercado Libre, encontramos

que el precio en pesos constantes aumentó para casi todos los tipos de inmuebles pre y post Covid-19, [en parte debido a una nueva ley de alquileres vigente desde el 1 de julio](#). Se destaca una variación en el precio por metro cuadrado descubierto entre ambos años, aumentando en promedio 8 % para los monoambientes, 11 % para los departamentos con 2 ambientes y 43 % para aquellos con 6 o más ambientes.

Si bien el aumento del trabajo remoto es un hecho, para que pueda ser considerado un mecanismo que afecta la composición del precio de los alquileres cabe preguntarse si esta práctica es percibida como transitoria o permanente. Una encuesta realizada por el portal de empleos [Bumeran](#) sobre la nueva modalidad de trabajo durante la cuarentena en Argentina indica que el 78 % de 1.800 encuestados “opina que el trabajo remoto pasará a ser una modalidad habitual de trabajo dentro de las empresas”. Esto nos da indicios de que el trabajo remoto podría ser un posible mecanismo que explica el cambio visto en los datos.

El trabajo se organiza de la siguiente manera. La sección 2 describe la literatura previa. La sección 3 discute los datos y la sección 4 detalla la metodología a utilizar. Por último, la sección 5 presenta los resultados preliminares.

2. Literatura previa

Existe una amplia literatura cuyo interés es predecir y caracterizar a los principales determinantes de los precios de las viviendas, dada la relevancia de estos en las decisiones de consumo e inversión de los hogares y empresas. Antes de decidir la compra de una vivienda, los hogares deben considerar muchas más variables que en cualquier otra transacción. Deben decidir el tipo de vivienda, la forma en que será financiada, su ubicación y sus características acorde a sus preferencias.

[Fan, Ong y Koh 2006](#) realizan un modelo predictivo para los precios de venta de inmuebles en Singapur. Los autores utilizan un modelo predictivo de CART, basando la predicción en las características de las viviendas, tanto estructurales, del barrio y del tipo de esta. Lo que encuentran es que las variables que explican el precio de la vivienda son distintas dados distintos tamaños de la vivienda. En el caso de estudio, mientras que en los departamentos de menos de 3 ambientes, son variables como la cantidad de metros cuadrados o la antigüedad de la vivienda los principales predictores del precio, para los departamentos más grandes la calidad, los alrededores y los servicios de la vivienda se convierten en predictores más relevantes. Este tipo de no linealidades es la que justifica para los autores el uso de CART. Sin embargo, este enfoque también tiene problemas para lidiar con variables que son relevantes en todas las observaciones. A su vez, como la variable a predecir es continua, CART presenta inconvenientes para poder predecir un único índice de

precios.

Una estrategia similar siguieron autores como [Özsoy y Şahin 2009](#), que utilizaron también CART para modelar los determinantes principales de los precios de las viviendas en Estambul. Ellos encuentran al tamaño, la seguridad y la ubicación como los principales determinantes de los precios siguiendo esta metodología. La elección de CART nuevamente se basa en que permite seleccionar variables y las interacciones no lineales que ocurren entre ellas. De esta manera en el caso de la determinación de los precios, los autores encuentran que a ciertos tamaños de la vivienda distintos factores resultan más relevantes.

Otros autores que plantearon la predicción de un índice de precios encontraron otras estrategias como óptimas, por ejemplo [Khalafallah 2008](#) utiliza un modelo de redes neuronales para predecir un índice de precios inmobiliarios para Estados Unidos en el contexto de la crisis hipotecaria. En un trabajo más reciente, [Park y Bae 2015](#) intentan predecir la evolución de un índice de precios inmobiliarios para Virginia, Estados Unidos y encuentran que el algoritmo de RIPPER es el que arroja mejores predicciones, por encima de otros métodos como redes neuronales, Support Vector Machine o AdaBoost.

Nuestro trabajo se enmarca dentro de esta literatura, planteando como aporte determinar los principales predictores del precio de la vivienda antes y después del impacto del Covid-19 y las medidas que se tomaron para enfrentarlo.

El impacto del Covid-19

El sector inmobiliario se vio golpeado como el conjunto de la economía a nivel global por la irrupción del virus del Covid-19. A partir del fuerte cambio en la vida cotidiana que las medidas para enfrentar la pandemia significaron, se observó un fenómeno de una fuerte caída en la demanda de viviendas en barrios con alta densidad de población. Este fenómeno fue aún mas intenso en aquellos distritos particularmente golpeados por el virus. Es particularmente interesante que este descenso en la demanda relativa de viviendas en sectores de alta densidad de población se mantuvo incluso luego de que se relajaran las medidas más estrictas de confinamiento y que el virus disminuyera su propagación ([Sitian y Yichen 2020](#)). Esto puede ser parcialmente explicado por la disminuida necesidad de vivir cerca del empleo gracias al trabajo remoto.

Realizando una primera exploración de los precios de alquileres de las viviendas en Mercado Libre para el AMBA antes y después de la pandemia del Covid-19 encontramos un cambio de precios relativos a favor de los espacios verdes y abiertos. Consideramos como posible hipótesis para explicar esta evolución, en congruencia con la reciente literatura que analiza el impacto del Covid-19 sobre el mercado inmobiliario, a una modificación de las preferencias de la demanda.

En particular, evaluamos la posibilidad de que haya ocurrido un cambio de expectativas acerca de la evolución futura del trabajo remoto. Por lo tanto, esto implicaría un mayor tiempo que los individuos podrían pasar en el hogar. En este sentido, nuestro trabajo podría constituir tanto un aporte a la literatura reciente que intenta analizar el impacto del Covid-19 sobre el sector inmobiliario, como así también a la literatura que trata sobre la evolución del trabajo a distancia.

Las nuevas tecnologías de la comunicación posibilitaron un aumento sostenido del trabajo remoto durante la última década. [Una encuesta de Gallup](#) del año 2016 para los Estados Unidos arrojaba que al menos un 43 % de la fuerza de trabajo de aquel país había realizado trabajo a distancia al menos alguna vez. Mas aún, [una encuesta del año 2018](#) llevada a cabo por la empresa IWG indicaba que dentro del ámbito profesional en Europa alrededor de un 70 % declaraba realizar trabajo a distancia por lo menos un día a la semana ([Cook 2019](#)).

Esta modalidad de trabajo podría ser beneficiosa tanto para empleados como empleadores al permitirles a estos conseguir trabajadores calificados que no necesariamente deban vivir en las inmediaciones de donde la compañía este ubicada. [Felstead y Henseke 2017](#) afirman, basados en evidencia de Reino Unido, que la tendencia al aumento del trabajo remoto es un hecho fundamental del siglo XXI. Encuentran además que el aumento en el mismo está relacionado con un mayor compromiso y satisfacción con el trabajo, aunque al costo de una mayor intensificación del ritmo de trabajo y de una incapacidad agrandada de separar el mundo laboral y personal. En la misma línea, [Bloom et al. 2014](#) aportan evidencia sobre el aumento en la productividad y la satisfacción laboral del trabajo desde casa, en base a un experimento con empleados de call center en China.

3. Descripción de la base de datos de Mercado Libre

La base de datos es provista por Mercado Libre, empresa dedicada mayormente a proveer una plataforma web para compra y ventas de bienes, tanto nuevos como usados. En los últimos años Mercado Libre también incursionó en el mercado de propiedades (tanto alquiler como compra y venta). En la página web de Mercado Libre, los usuarios registrados pueden publicar avisos de alquiler o de venta de inmuebles, estos avisos pueden ser visitados por cualquiera que ingrese el sitio web, en caso de estar interesado en el inmueble puede contactarse con el autor de la publicación.

Los datos provistos incluyen variables como el precio de alquiler del inmueble, la fecha en la que fue publicado y la fecha en la cual se actualizó algún dato de la publicación. Cantidad de ambientes, si cuenta con pileta, gimnasio, salón de fiestas, salón de usos múltiples, etc.

Limpieza de base de datos

La base de datos atraviesa un proceso de limpieza, en primer lugar se quitan publicaciones con precio y metros cuadrados iguales a 0. También se remueven publicaciones que no especifican de qué tipo de inmueble se trata.¹ Solamente se seleccionan publicaciones que hacen referencia a alquileres de departamentos.

La base de datos cuenta con muchas observaciones por una misma publicación, se repiten cada vez que la publicación se actualiza. Para lidiar con este problema se utilizó la fecha de la última actualización por cada publicación.

Por la presencia de outliers² es necesario realizar cierto ajuste en la base de datos. En [IMF 2020](#), pág. 21 se recomienda lidiar con outliers y errores a través de definir primero un rango de valores plausibles y posteriormente eliminar aquellas observaciones que escapan de estos valores. Para esto se quitaron todas las observaciones que estén por arriba del percentil 95 en términos de metros cuadrados, precio y metros cuadrados descubiertos, por otro lado también se quitaron todas las observaciones que estén por debajo del percentil 5 en términos de precio, metros cuadrados y metros cuadrados descubiertos.

Creación de variables

Se crea la variable de metros descubiertos con la diferencia de metros cuadrados y metros cuadrados cubiertos. También se creará posteriormente una variable que mide qué tan céntrica es la zona en la que se encuentra el inmueble. Para inmuebles dentro de CABA se toma el centroide del barrio y se computará la distancia euclidiana hacia el obelisco, de esta manera se tendrá una aproximación de cercanía al centro para cada propiedad en base al barrio donde se ubica. Para computar los precios reales se toma como mes base Enero de 2018 y se utiliza el aumento de índice de precios al consumidor publicados a través del INDEC para deflacionar la serie de precios. Con excepción de noviembre 2020 donde en vez de utilizar el IPC se utiliza el relevamiento de expectativas de mercado (REM) publicado por el BCRA.

¹La existencia de valores faltantes puede tratarse a través de diferentes métodos, véase [Gelman e Hill 2006](#), cap. 25. En este caso por simplicidad se decide eliminar aquellas observaciones con valores faltantes en esta variable.

²Que probablemente surjan de la imputación manual de ciertas variables. Es posible que estos outliers surjan por error o intencionalmente. Por ejemplo si un usuario busca que su publicación aparezca entre las más baratas podría poner un precio de 1, de esta forma aparecerá primera aunque no sea el precio finalmente pactado en un contrato.

Estadísticas descriptivas

En la [Tabla 1](#) se pueden observar ciertas características de la base de datos, la misma se construye una vez limpiada esta y eliminado los valores outliers como se explicó en la sección anterior.

Tabla 1: Estadísticas descriptivas

	Media	Mín	Mediana	Max
M^2 totales	58.47	27	52	159
Precio real M^2	157.71	11.01	151.82	650.44
Precio real M^2 descubierto	1,876.18	60.25	1,573.62	20,274.82
Precio real	8,826.57	1,751.01	8,050.30	20,366.28
M^2 descubierto	7.59	1	5	35
Cantidad de observaciones: 73,344				

Estos valores son calculados a partir de la base de datos completa una vez que se realiza la limpieza por percentiles descrita anteriormente. Sin embargo en lo que respecta a análisis posteriores el podado se realizará para la base de datos pre cuarentena y post implementación de cuarentena.

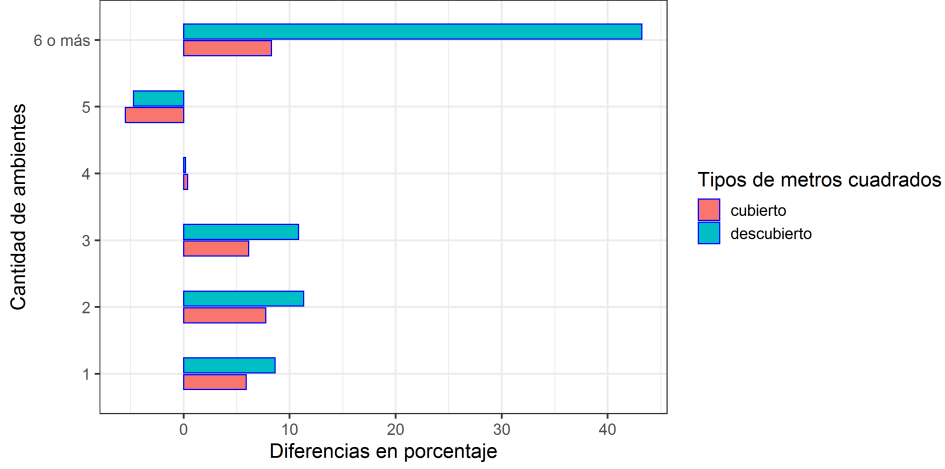
Por otro lado, una vez que la base de datos se divide pre cuarentena y post implementación de cuarentena y posteriormente se quitan outliers, nos encontramos con un total de 92,043 publicaciones pre cuarentena y 59,124 publicaciones en tiempos de cuarentena. En lo que respecta a correlaciones entre ciertas variables y el precio real del alquiler del departamento, podemos observar en la [Tabla 2](#) que tanto antes de la cuarentena como después de la misma hay una correlación positiva entre precio real y características como tener gimnasio, pileta, salón de fiestas, etc. Sin embargo, posterior a la cuarentena la correlación es menor, esto podría estar sugiriendo que ahora las zonas comunes no logran aumentar tanto el precio. El efecto de corto plazo que tiene la pandemia sobre los precios de inmuebles es coherente con este resultado.

Tabla 2: Coeficientes de correlación entre precios constantes y diferentes características de departamentos

Período	<i>Gimnasio</i>	<i>Pileta</i>	<i>Salon de fiestas</i>	<i>SUM</i>
Pre cuarentena	0.27	0.31	0.11	0.24
Curantena	0.19	0.25	0.09	0.18

En la [Figura 1](#) están graficadas las diferencias porcentuales de precios constantes por metros cuadrados entre pre cuarentena y post cuarentena. En la mayoría de los casos el aumento es mayor en el caso de precio por metros cuadrados descubiertos. Para el caso de 5 ambientes, se puede ver una reducción del precio, pero para los metros cuadrados descubiertos la reducción de precios es menor. Esto indicaría que espacios verdes o abiertos como balcones o terrazas son más cotizados en tiempos de cuarentena que antes.

Figura 1: Diferencias en el aumento de precios por metros cuadrados



4. Metodología

En los apartados anteriores se mencionó que varios trabajos buscan obtener el modelo que mejor prediga en términos de error cuadrático medio. En este trabajo a diferencia de otros se busca analizar cualitativamente cuáles son los mejores predictores antes de la cuarentena en comparación a durante la cuarentena. En este sentido no se utilizará una pila de modelos sino simplemente CART y LASSO. A partir de estos dos modelos se puede analizar la importancia de los determinantes de precios en base a los primeros nodos en el caso de CART y tamaño de coeficientes en el caso de LASSO.

CART

En primer lugar a través de CART se podrán captar no linealidades en los predictores de precios de las publicaciones. Por otro lado, es posible detectar cuáles son predictores relevantes teniendo en cuenta cuáles variables aparecen en nodos más cercanos al origen del árbol. Por ejemplo, en caso de que la cantidad de ambientes sea muy relevante, el primer nodo se bifurcará en otros dos por la cantidad de ambientes que tenga un departamento. CART utiliza un algoritmo para determinar qué variable X_j utilizará en cada partición y qué punto s de esa variable se utilizará, y luego itera este mismo algoritmo.

$$\min_{j,s} \left[\min_{c_1} \sum_{X_i \in R_1(j,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{X_i \in R_2(j,s)} (y_i - c_2)^2 \right]$$

Para evitar el sobreajuste a los datos, el método de CART va a ir reduciendo el tamaño del árbol para satisfacer un criterio de capacidad predictiva fuera de la muestra. Este método puede ser muy útil para representar las variables más relevantes a la hora de determinar el precio de los alquileres de las viviendas. Sin embargo, en presencia de una estructura lineal de los datos, con variables que sean determinantes para toda la muestra y no solo dentro de cada partición, CART podría ser un método poco robusto, donde pequeñas alteraciones en los datos podrían producir alteraciones en los predictores.

Por este motivo, métodos como Random Forest, con el caso particular de Bagging, pueden servir para reducir la varianza. Estos pueden utilizarse con ayuda del paquete `rpart` en R. La idea detrás es realizar varias muestras bootstrap y tomar como predictor al promedio de las predicciones realizadas con cada una de las muestras.

LASSO

En segundo lugar se utilizará LASSO para captar cuáles son los predictores más relevantes en base al tamaño de los coeficientes asociados.

$$\hat{\beta}_{lasso} = \min_{\beta} \sum_{n=1}^N \frac{1}{2} (y_n - \beta x_n)^2 + \lambda \sum_{i=1}^p |\beta_i|$$

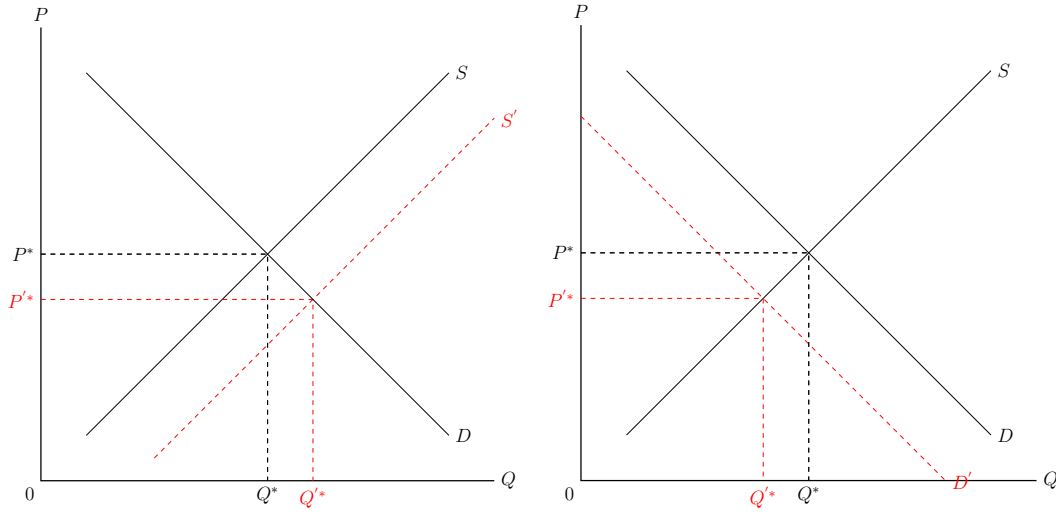
LASSO cumple tanto un rol de reducción de dimensionalidad (por eliminar coeficientes) como de mejorar las predicciones a costa de producir estimaciones sesgadas. Al momento de utilizar LASSO es necesario tomar una decisión sobre el λ óptimo, este se seleccionará a través de K-fold cross validation. En general, se recomienda utilizar 5-fold cross validation ([Breiman y Spector 1992](#)) o 10-fold cross validation ([Kohavi et al. 1995](#)). El método de Cross validation dividirá la muestra en K partes y utilizará todas menos una parte para estimar el modelo y esa parte restante se utilizará para testear el modelo. Este procedimiento se realizará de tal forma que todas las partes hayan cumplido un rol de entrenamiento y de testeo.

5. Resultados preliminares

En un apartado anterior se mencionó que hay cambios de precios constantes. Estos cambios en los precios pueden deberse a diferentes causas, en principio es imposible distinguir en un contexto de este tipo si los cambios son producidos por un desplazamiento en la oferta o en la demanda (véase por ejemplo la [Figura 2](#) dos casos totalmente diferentes pero con mismo resultados en términos de precios). No obstante, como se dijo también anteriormente es posible generar hipótesis

que den cuenta de los resultados obtenidos.

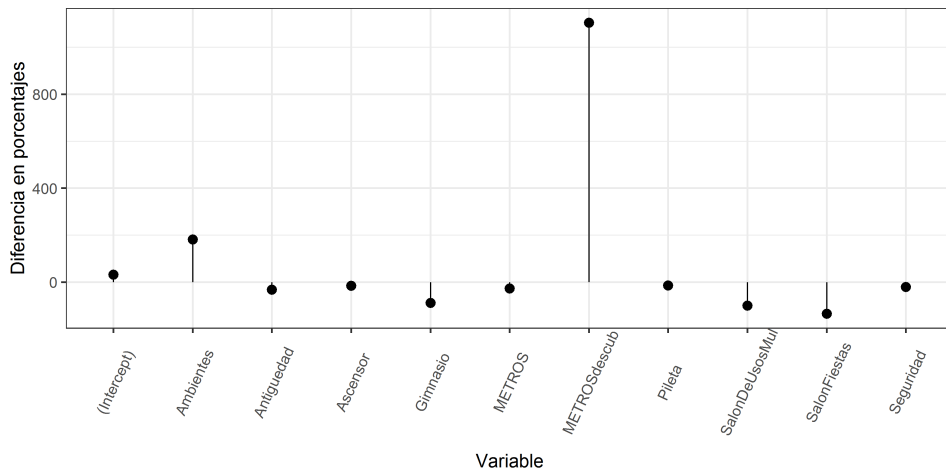
Figura 2: Dos movimientos que llevan al mismo resultado en términos de precios de equilibrio



LASSO

En lo que respecta a LASSO, podemos ver que en tiempos de cuarentena los coeficientes cambian (Figura 3), sobretodo el coeficiente asociado a metros cuadrados descubiertos. Este resultado indica luego de la irrupción del Covid-19, la cantidad de metros cuadrados descubiertos que tiene un departamento, pasa a ser un predictor muy relevante del precio de alquiler. Variables como Salón de Fiestas, Salón de Usos múltiples y Gimnasio, son menos importantes en tiempos de cuarentena, lo cual es coherente con el efecto de corto plazo antes mencionado. Para el caso de Piletas, esto no se cumple. También puede observarse que ahora la cantidad de ambientes es una variable más relevante una vez que se implementa la cuarentena.

Figura 3: Diferencias de coeficientes a través de LASSO



Referencias

- Bloom, Nicholas et al. (nov. de 2014). “Does Working from Home Work? Evidence from a Chinese Experiment *”. En: *The Quarterly Journal of Economics* 130.1, págs. 165-218. ISSN: 0033-5533. DOI: [10.1093/qje/qju032](https://doi.org/10.1093/qje/qju032). eprint: <https://academic.oup.com/qje/article-pdf/130/1/165/30629971/qju032.pdf>. URL: <https://doi.org/10.1093/qje/qju032>.
- Breiman, Leo y Philip Spector (1992). “Submodel selection and evaluation in regression. The X-random case”. En: *International statistical review/revue internationale de Statistique*, págs. 291-319.
- Cook, Sarah (2019). “BENEFITS OF REMOTE WORKING”. En: *Making a Success of Managing and Working Remotely*. IT Governance Publishing, págs. 12-17. ISBN: 9781787781283. URL: <http://www.jstor.org/stable/j.ctvkjb2pv.7>.
- Fan, Gang-Zhi, Seow Eng Ong y Hian Chye Koh (2006). “Determinants of House Price: A Decision Tree Approach”. En: *Urban Studies* 43.12, págs. 2301-2315. DOI: [10.1080/00420980600990928](https://doi.org/10.1080/00420980600990928). eprint: <https://doi.org/10.1080/00420980600990928>. URL: <https://doi.org/10.1080/00420980600990928>.
- Felstead, Alan y Golo Henseke (2017). “Assessing the growth of remote working and its consequences for effort, well-being and work-life balance.” En: *New Technology, Work Employment* 32.3, págs. 195-212. ISSN: 02681072. URL: <https://eza.udes.edu.ar/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=iih&AN=126406313&lang=es&site=ehost-live>.
- Gelman, Andrew y Jennifer Hill (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.
- IMF (2020). *Residential property price index: Practical compilation Guide*. International monetary fund.
- Khalafallah, A. (2008). “Neural network based model for predicting housing market performance”. En: *Tsinghua Science and Technology* 13.S1, págs. 325-328. DOI: [10.1016/S1007-0214\(08\)70169-X](https://doi.org/10.1016/S1007-0214(08)70169-X).
- Kohavi, Ron et al. (1995). “A study of cross-validation and bootstrap for accuracy estimation and model selection”. En: *Ijcai*. Vol. 14. 2. Montreal, Canada, págs. 1137-1145.
- Özsoy, Onur y Hasan Şahin (mayo de 2009). “Housing price determinants in Istanbul, Turkey: An application of the classification and regression tree model”. En: *International Journal of Housing Markets and Analysis* 2, págs. 167-178. DOI: [10.1108/17538270910963090](https://doi.org/10.1108/17538270910963090).
- Park, Byeonghwa y Jae Kwon Bae (2015). “Using machine learning algorithms for housing price prediction: The case of Fairfax County, Virginia housing data”. En: *Expert Systems with Applications* 42.6, págs. 2928-2934. ISSN: 0957-4174. DOI: <https://doi.org/10.1016/j.eswa.2014.11.040>. URL: <http://www.sciencedirect.com/science/article/pii/S0957417414007325>.
- Sitian, Liu y Su Yichen (2020). “The Impact of the COVID-19 Pandemic on the Demand for Density: Evidence from the U.S. Housing Market”. En: *Social Science Research Network*.