

Data Science 2022

TRABAJO PRÁCTICO II

Estimación del precio por m² para departamentos en CABA

Grupo 4

Atienza Rela, Guadalupe

Atienza Rela, Macarena

Carrero, Luis

Federico, Santiago Raul

Roa Herrera, Juan

1

Selección de los datos a modelar

2

Modelización

3

Evaluación de los modelos y conclusiones

Selección de los datos a modelar

Selección de los datos a modelar

FILAS

Registros por tipo de propiedad (deptos) y ubicación (CABA)

Registros duplicados

Registros con campos sin datos

Registros con valores atípicos muy extremos

Selección de
observaciones

COLUMNAS

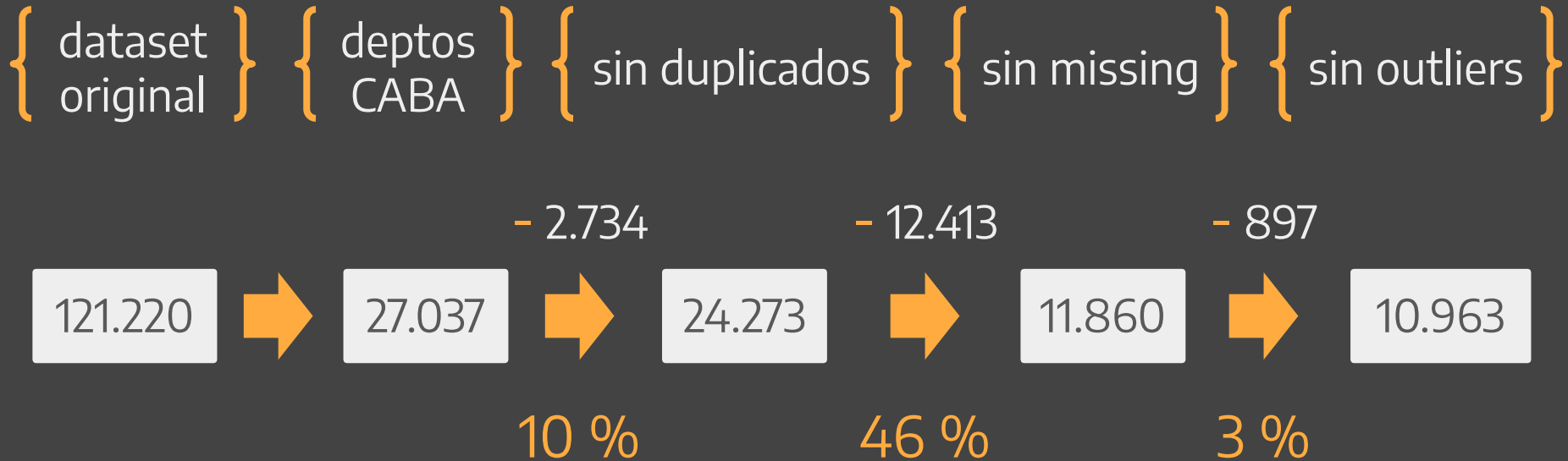
Variables de localización

Variables relativas a las características de las propiedades

Variables que no se van a utilizar

Selección y
creación de
variables

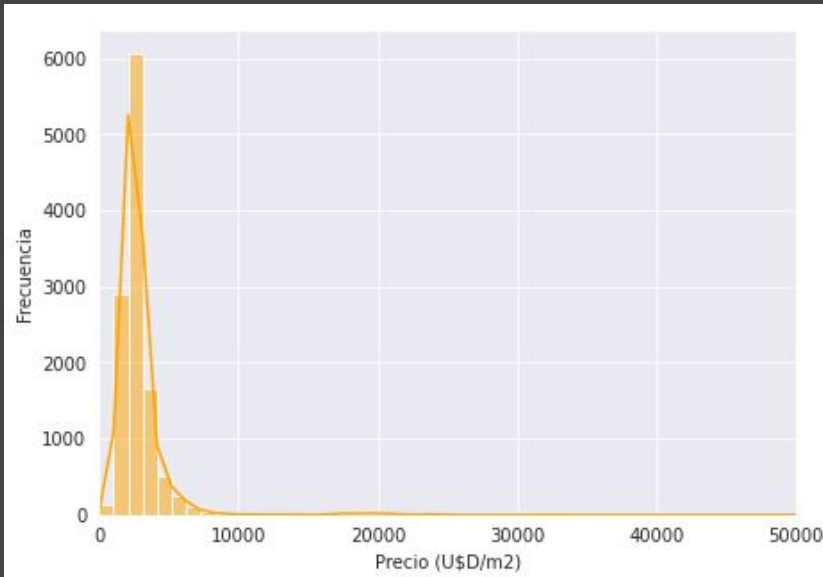
Selección de los datos a modelar: observaciones



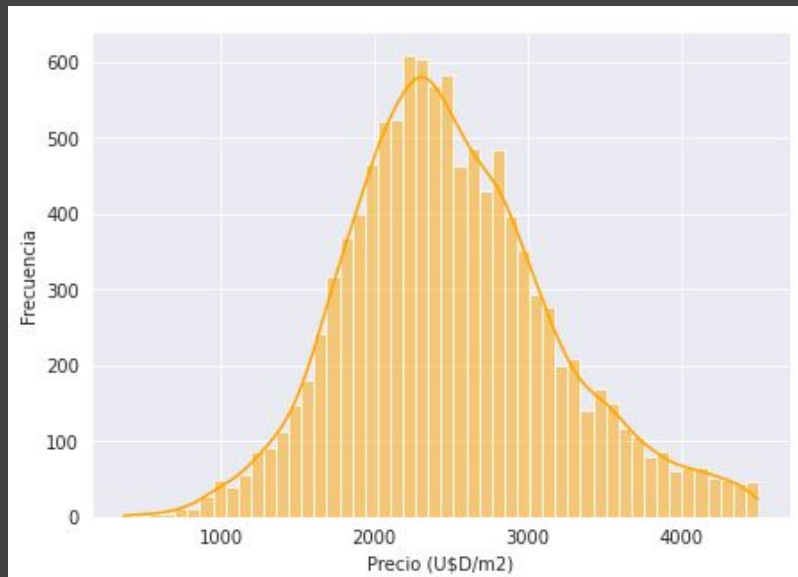
Selección de los datos a modelar: observaciones

Distribución de frecuencia de la variable Precio antes y después de eliminar los outliers

ANTES

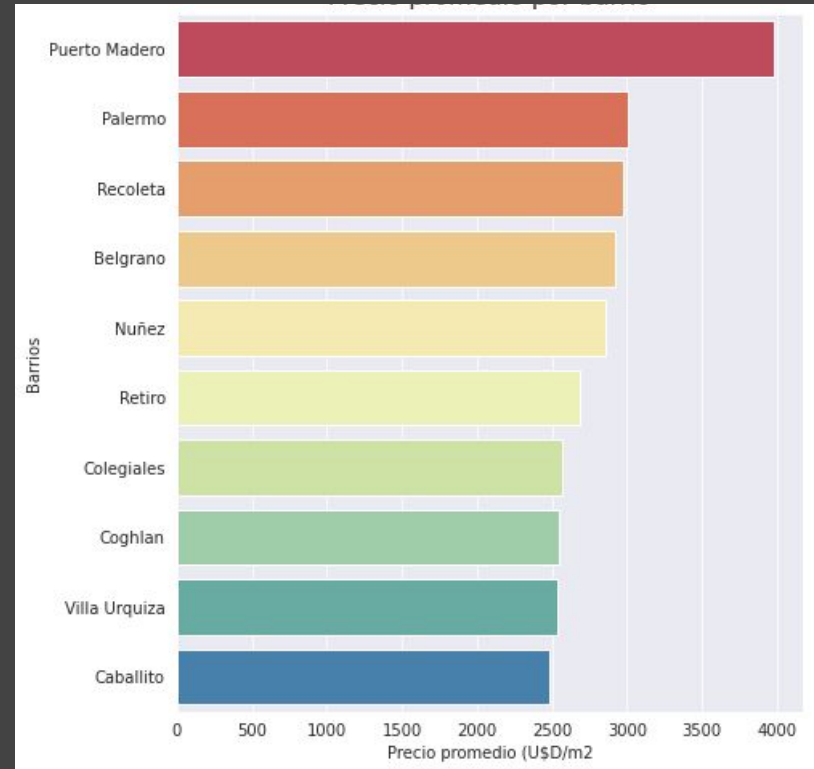


DESPUÉS



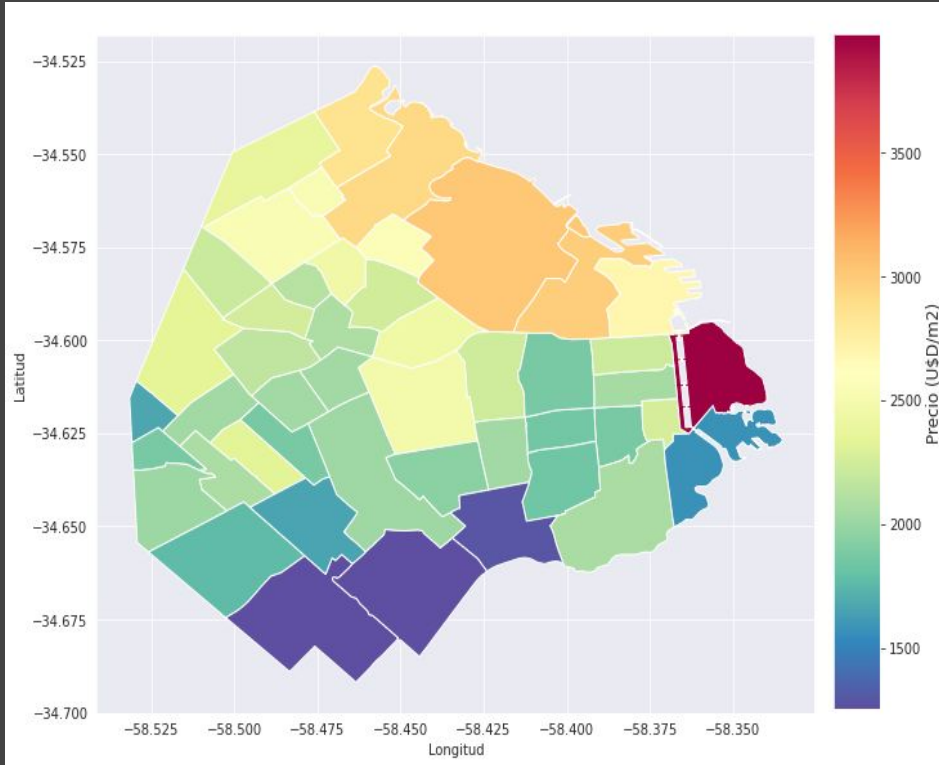
Selección de los datos a modelar: variables

barrio	Precio (USD por m2)			
	min	mediana	promedio	max
Puerto Madero	2964.91	4090.85	3978.06	4450.00
Palermo	895.77	2979.80	3010.90	4500.00
Recoleta	1275.00	2918.19	2977.50	4500.00
Belgrano	569.95	2890.91	2919.63	4422.54
Nuñez	1260.62	2819.67	2855.26	4500.00
Retiro	1424.24	2687.50	2690.07	4363.64
Colegiales	638.30	2504.10	2571.23	4126.98
Coghlan	516.43	2560.98	2550.76	4166.67
Villa Urquiza	697.67	2527.26	2534.81	4437.50
Caballito	868.85	2436.24	2481.29	4271.74



Selección de los datos a modelar: variables

MAPA DE PRECIOS POR BARRIO



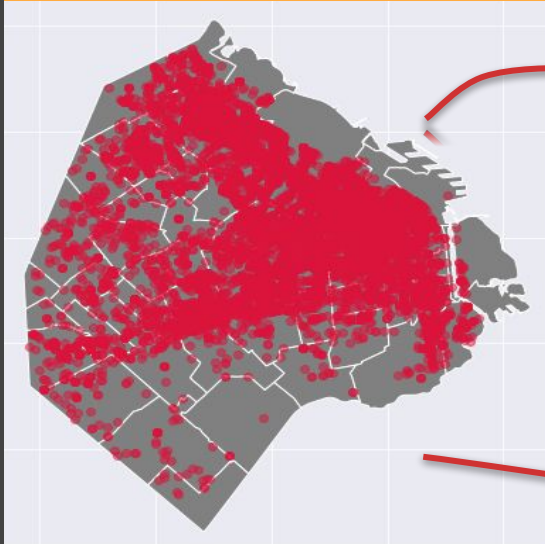
NUEVAS VARIABLES

- Ubicación en el eje norte / sur
- Ubicación en el eje este / oeste
- Cercanía con el subte
- Cercanía con una zona comercial
- Cercanía con un barrio informal

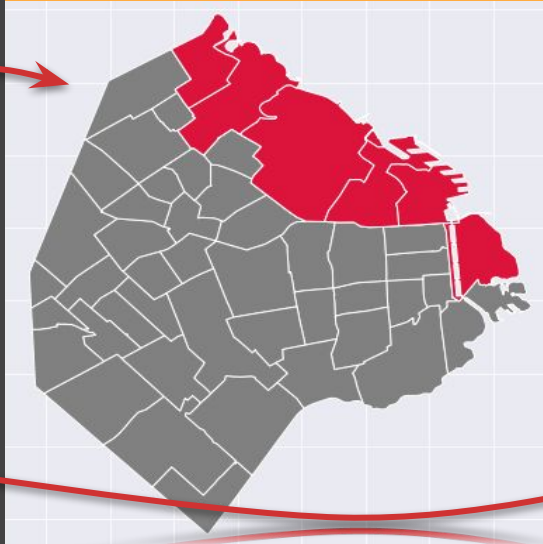
{ para poder captar los efectos de la localización en el modelo }

Selección de los datos a modelar: variables

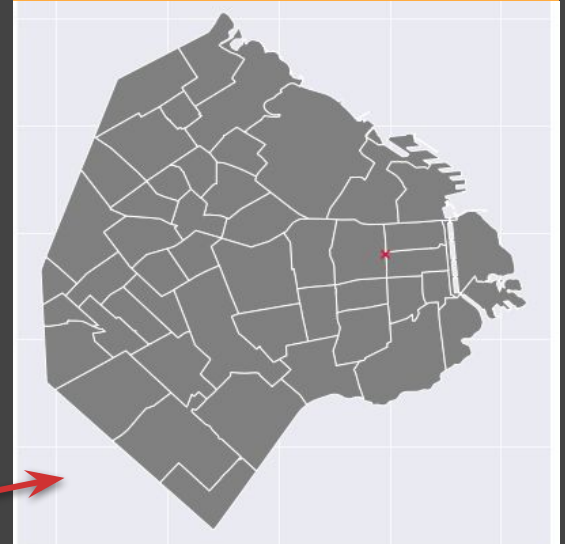
UBICACIÓN DEPARTAMENTOS



EJE NORTE / SUR



EJE ESTE / OESTE



{ Creación de una variable dummy que identifica las propiedades que se encuentran en barrios del eje norte }

{ Creación de una variable continua que mide la distancia de las propiedades al centro urbano }

Selección de los datos a modelar

LOCALIZACIÓN	CARACTERÍSTICAS PROPIAS
DU Ubicación en barrios del eje norte	CO Superficie total
CO Distancia del centro	CO Superficie cubierta
DU Cercanía con el subte	CO Cantidad de ambientes
DU Cercanía con una zona comercial	DU Con amenities
DU Cercanía con un barrio informal	DU A estrenar



Precio (U\$D)
por m2

{ variables predictoras o *features* }

{ variable a predecir o *target* }

Modelización, evaluación de
los modelos y conclusiones

Modelización

(1) { Regresión Lineal Simple }

(2) { Normalización }

(3) { Train, test split }



MODELOS

Regresión Lineal Múltiple

Regresión Ridge

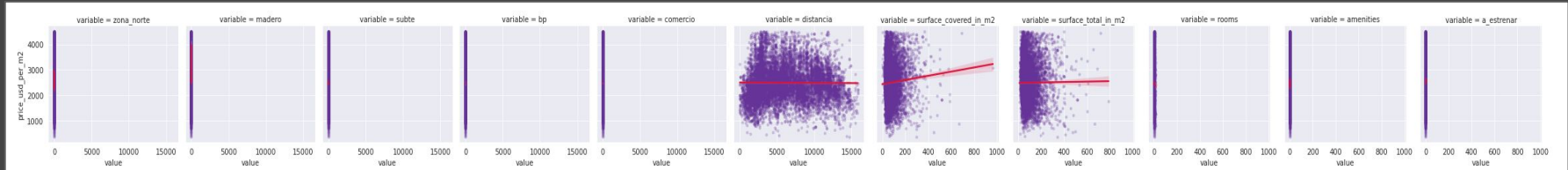
Regresión Lasso

(4)

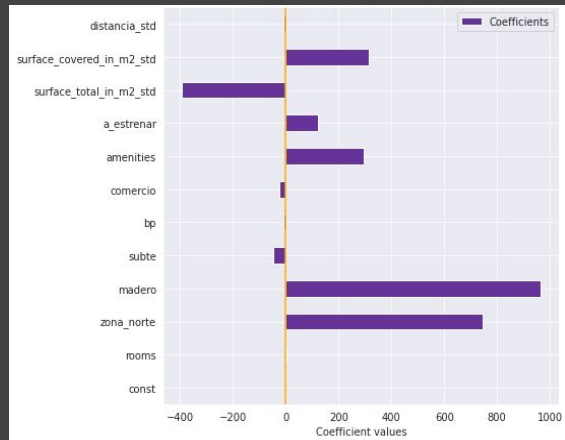
(5)

(6)

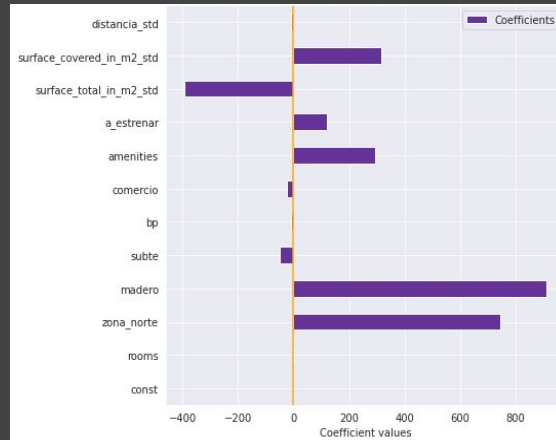
Modelización: regresión lineal simple, múltiple y regularización



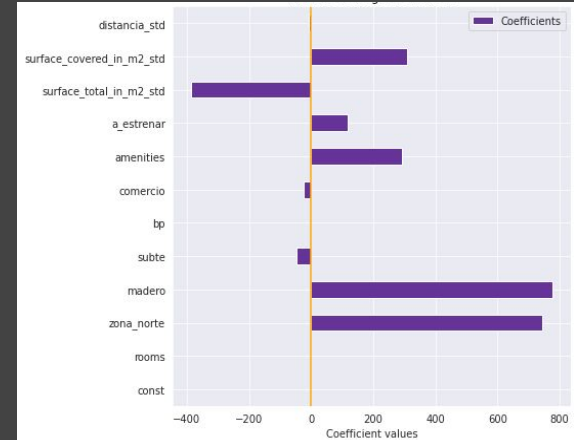
Regresión Lineal Múltiple



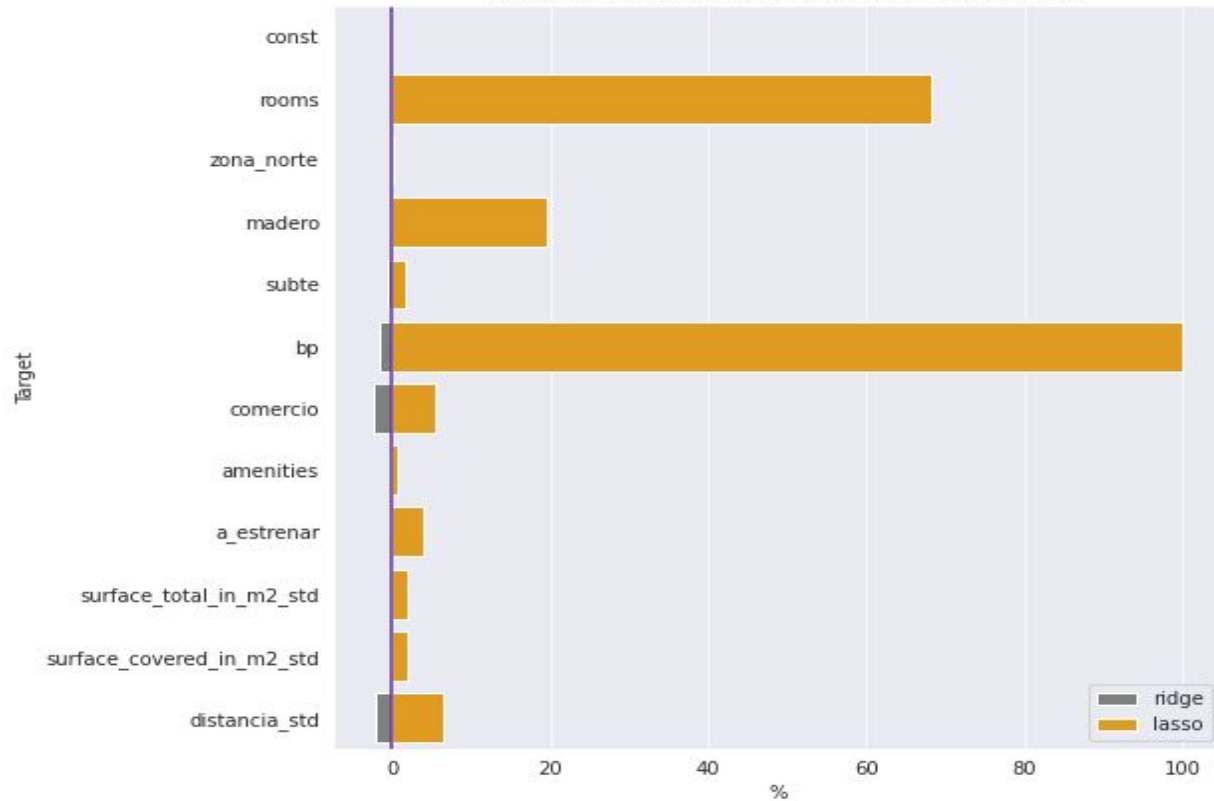
Regresión Ridge



Regresión Lasso



Modelización: variación respecto al modelo inicial



para ver notebook en GitHub click [aquí](#)

¡Muchas Gracias!