



Trabajo Práctico 2

Modelado con Regresión Lineal

Grupo 6 - Feb 2024



PREPARACIÓN DEL DATASET

DEFINICIÓN DE COLUMNAS A UTILIZAR

- Seleccionamos las columnas significativas para el modelado.
- Seleccionamos y Target = 'PRICE_USD_PER_M2'
- Creación de nuevas columnas útiles para diferenciar el precio de las propiedades

DEFINICIÓN DE DATOS A UTILIZAR

- Seleccionamos a los **APARTAMENTOS DEL BARRIO DE PALERMO** como nuestra información base para realizar el modelado → Objetivo: Reducir la variabilidad del modelo al tener datos más homogéneos.
- Aplicamos filtros de criterio para eliminar Outliers
 - Sup Total entre 50 y 500 m2
 - Rooms entre 1 y 10



ARMADO DE MATRIZ DE FEATURES Y TARGET

COLUMNAS DUMMIES

- Se realizó la transformación de las columnas categóricas a columnas Dummies.

DEFINICIÓN DE X,Y

- Se armaron las matrices X y el vector Y.

NORMALIZACIÓN DE DATOS

- Estandarizamos las variables para eliminar sus distintas unidades de medida y evitar distorsiones debidas a diversas escalas utilizando el método `StandardScaler()`



SEPARACIÓN DE DATOS EN TRAIN, TEST

SELECCIÓN DE DATOS PARA TRAIN, TEST

- Utilizando el 70% de los datos para el train y 30% para test. Utilizamos el método `train_test_split`

REALIZAMOS LA PREDICCIÓN DE PRECIOS UTILIZANDO

- Regresión Lineal Múltiple
- Ridge
- Lasso

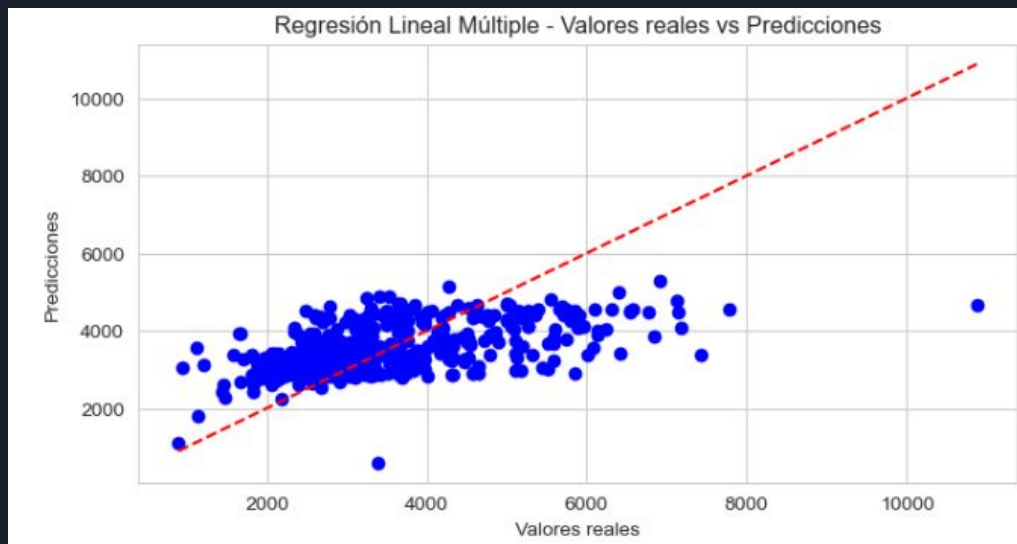
MATRIZ DE RESULTADOS

OBTUVIMOS LAS SIGUIENTES MÉTRICAS DE EVALUACIÓN

	Model	MSE	MAE	RMSE	R2	Alpha	Cant. Cols.
0	Regularización Ridge	1099979.03	809.41	1048.80	0.29	11.2	8.0
1	Regularización Lasso	1097218.54	807.96	1047.48	0.29	0.1	8.0
2	Regresión Lineal	1097238.22	807.96	1047.49	0.29	-	8.0

EXPLICACIÓN DE R^2

Un R^2 de aproximadamente 0.29 significa que alrededor del 29% de la variabilidad en la variable dependiente puede explicarse por las variables independientes incluidas en el modelo de regresión lineal múltiple. Esto sugiere que el modelo no ajusta bien los datos y deja una cantidad sustancial de variabilidad sin explicar. Es posible que se necesiten incluir más variables o utilizar un modelo más complejo para mejorar la capacidad predictiva del modelo.





Conclusiones

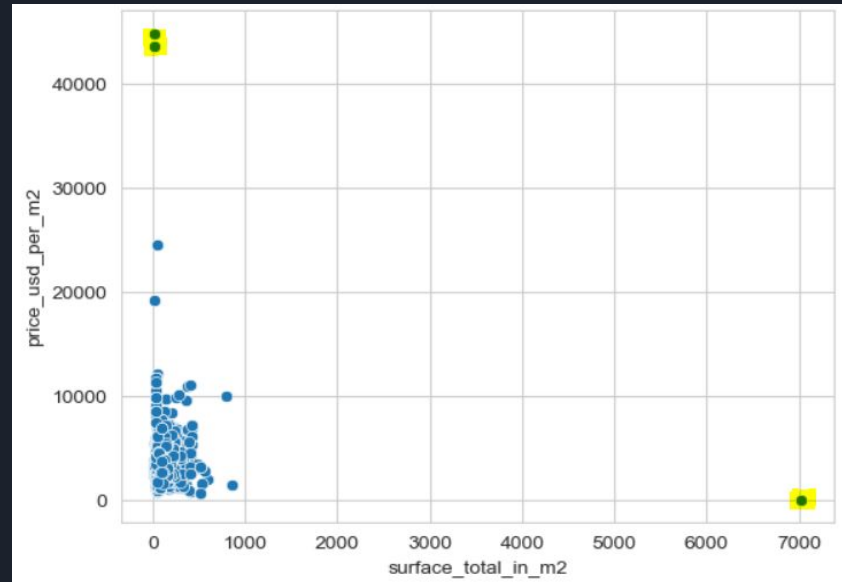
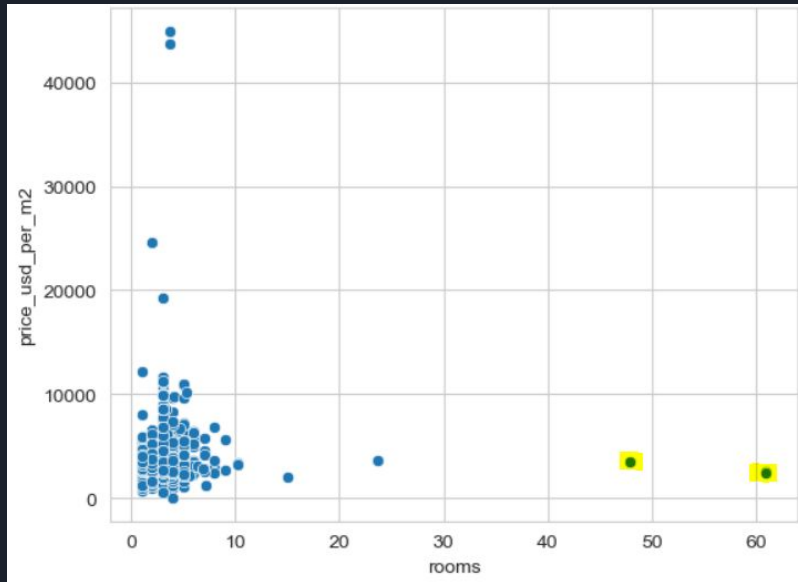
Conclusiones - **1**: Outliers

Conclusiones - **2**: Variabilidad

Conclusiones - **3**: Métricas y modelos

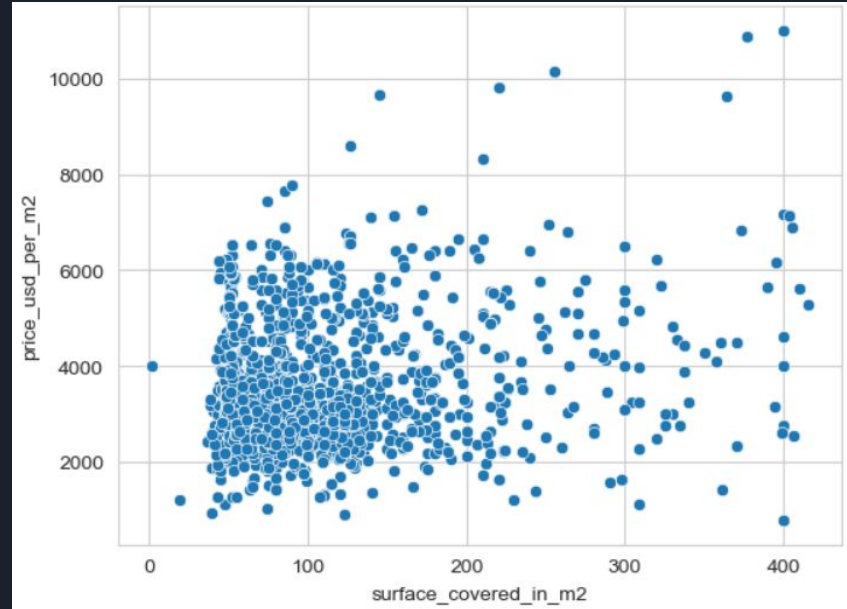
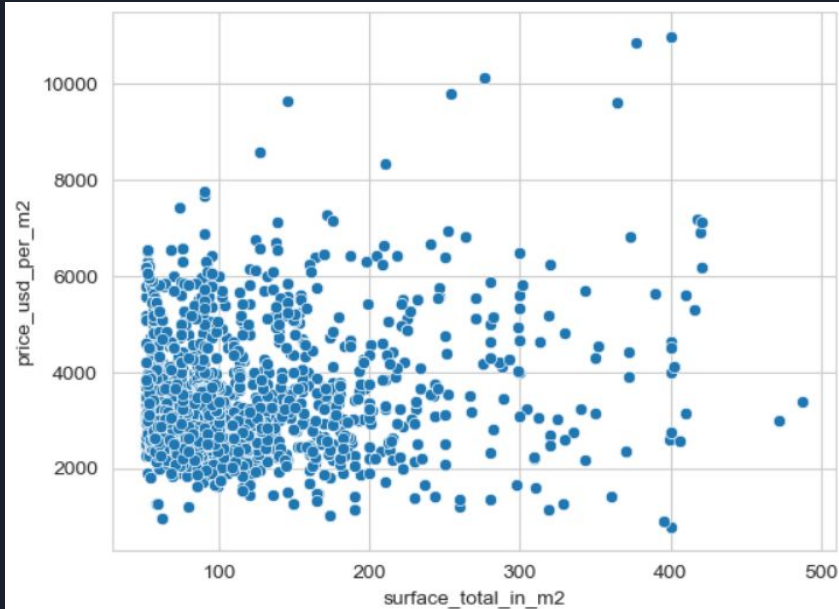
Conclusión - 1 Outliers

Presencia de outliers en las variables predictoras



Conclusión - 2 Variabilidad

Gran variabilidad de las predictoras con el target





Conclusión - 3 Métricas y modelo

Diferencia entre los modelos

Coefficientes de los modelos

	B0	surface_total_in_m2	surface_covered_in_m2	rooms	tiene_pileta_d	tiene_balcon_d	tiene_SUM_d	tiene_cochera_d	tiene_gimnasio_d
Ridge	2569.06	-7.54	9.83	56.12	371.97	96.74	138.36	356.06	629.54
Lasso	2553.26	-7.55	9.80	58.48	370.51	101.38	127.70	363.62	669.90
Regresion Multiple	2551.89	-7.55	9.80	58.73	370.52	102.08	127.98	363.85	670.31



Conclusión - 3 Métricas y modelo

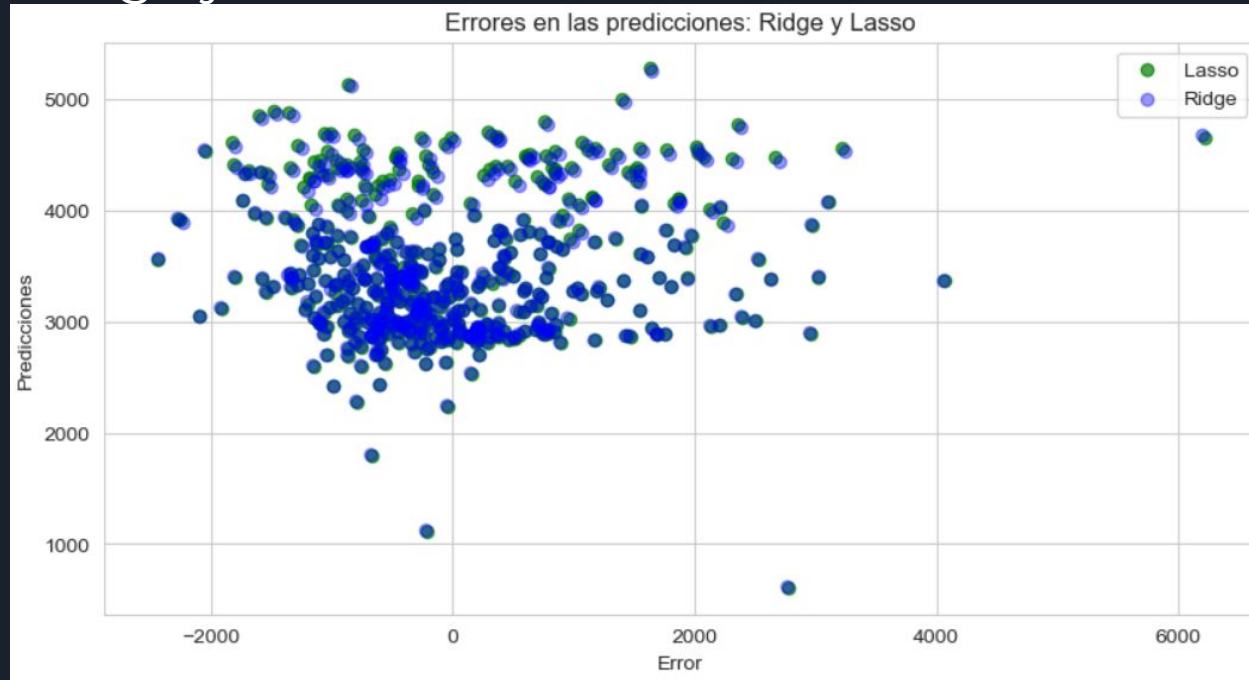
Diferencia entre los modelos

Métricas de los modelos

	Model	MSE	MAE	RMSE	R2	Alpha	Cant. Cols.
0	Regularización Ridge	1099979.03	809.41	1048.80	0.29	11.2	8.0
1	Regularización Lasso	1097218.54	807.96	1047.48	0.29	0.1	8.0
2	Regresión Lineal	1097238.22	807.96	1047.49	0.29	-	8.0

Conclusión - 3 Métricas y modelo

Ridge y lasso



Conclusión - 3 Métricas y modelo

Ridge y Regresión múltiple

