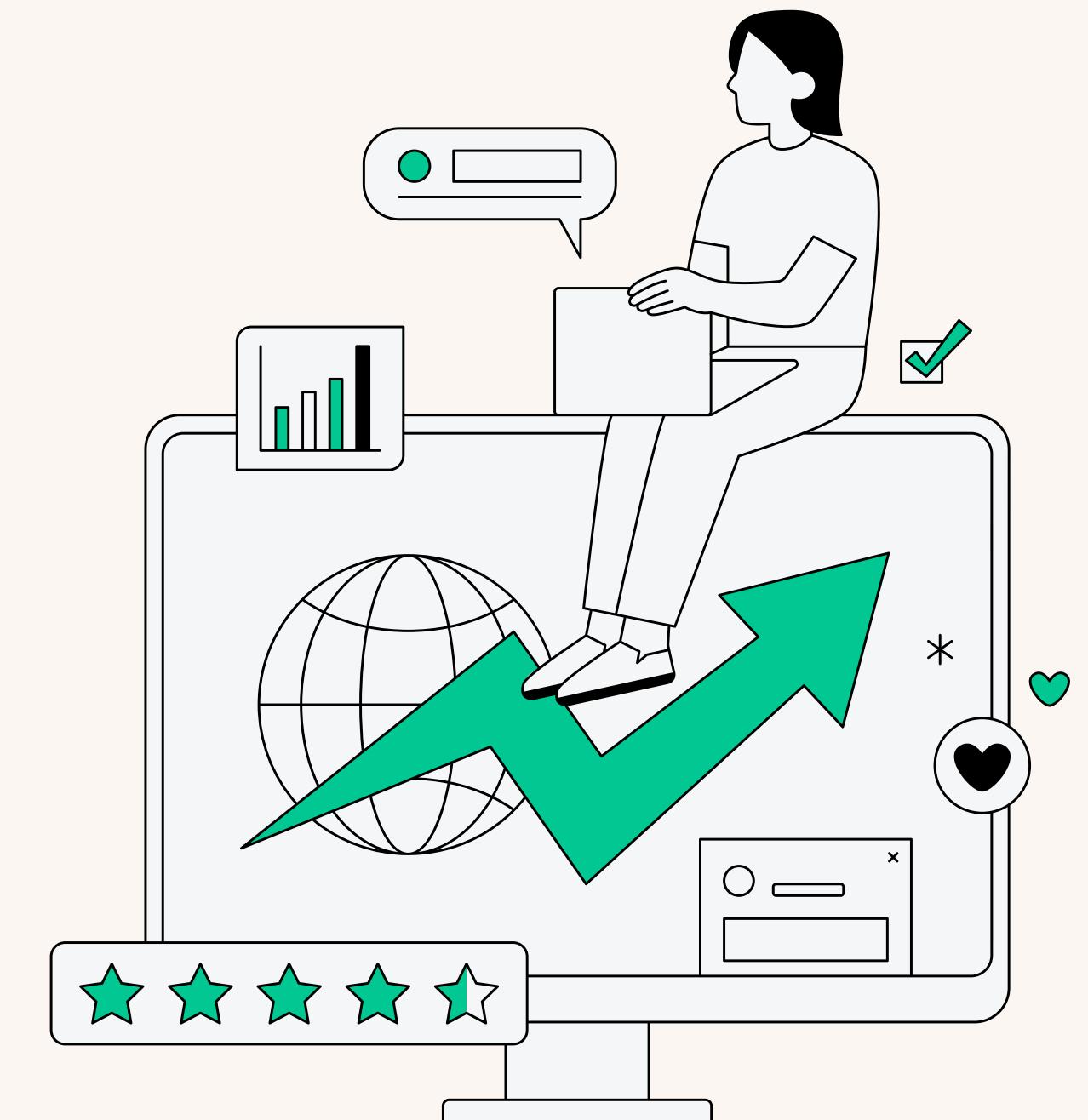


Universidad de los Andes

Analisis de resultados

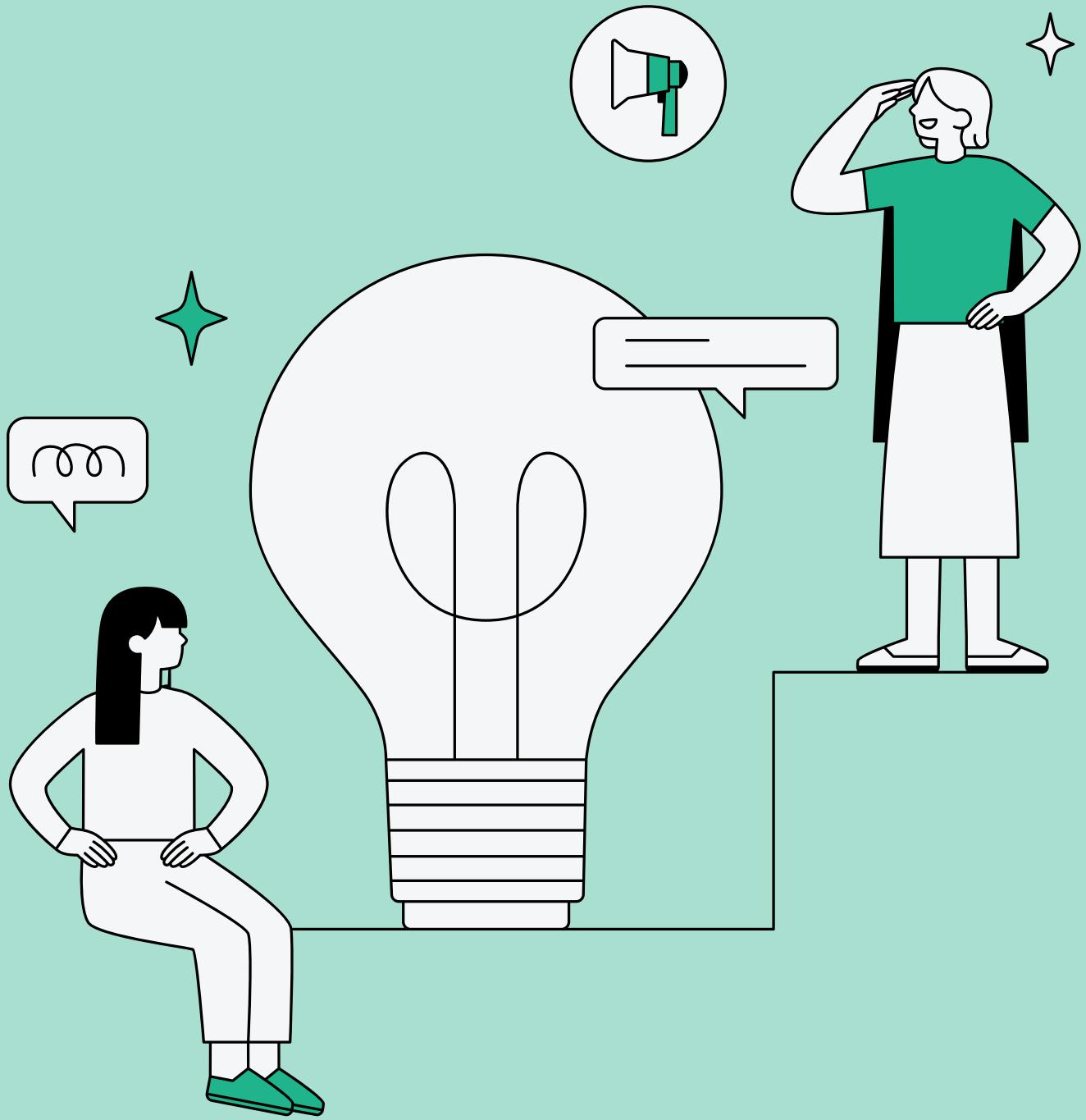
CompuAlpes

Ernesto José Duarte
Carlos Pinto Rodriguez
Juan Camilo Gonzales



Introducción

CompuAlpes, un líder en tecnología, se enfrenta al desafío de fijar precios precisos para sus portátiles en un mercado en constante cambio. Lo cual, esta presentación detallará nuestro proceso, desde la comprensión de datos hasta la evaluación de modelos, con el objetivo de respaldar decisiones de fijación de precios más informadas y estratégicas.



Tratamiento de datos

CONSISTENCIA:

Se identificaron valores con caracteres extraños en las variables “TypeName” y “Gpu_brand” donde se corrige la información eliminando los caracteres, por otro lado, se realiza la unión de los registros de valores únicos de los datos train que no tiene en los datos de test ya que tiene más valores únicos en varias categorías. Para valores desconocidos se clasifica en “Others”

UNICIDAD:

Se verificó la existencia de registros duplicados en el conjunto de datos, en la cual se encontraron 15 registros duplicados y se eliminaron estos registros duplicados para mantener la unicidad de los datos.

Tratamiento de datos

COMPLETITUD:

Se encontraron valores faltantes en las columnas 'Weight', 'Ppi', 'HDD', 'SSD', y Price. Por lo tanto, se eliminaron las entradas con valores vacíos o nulos.

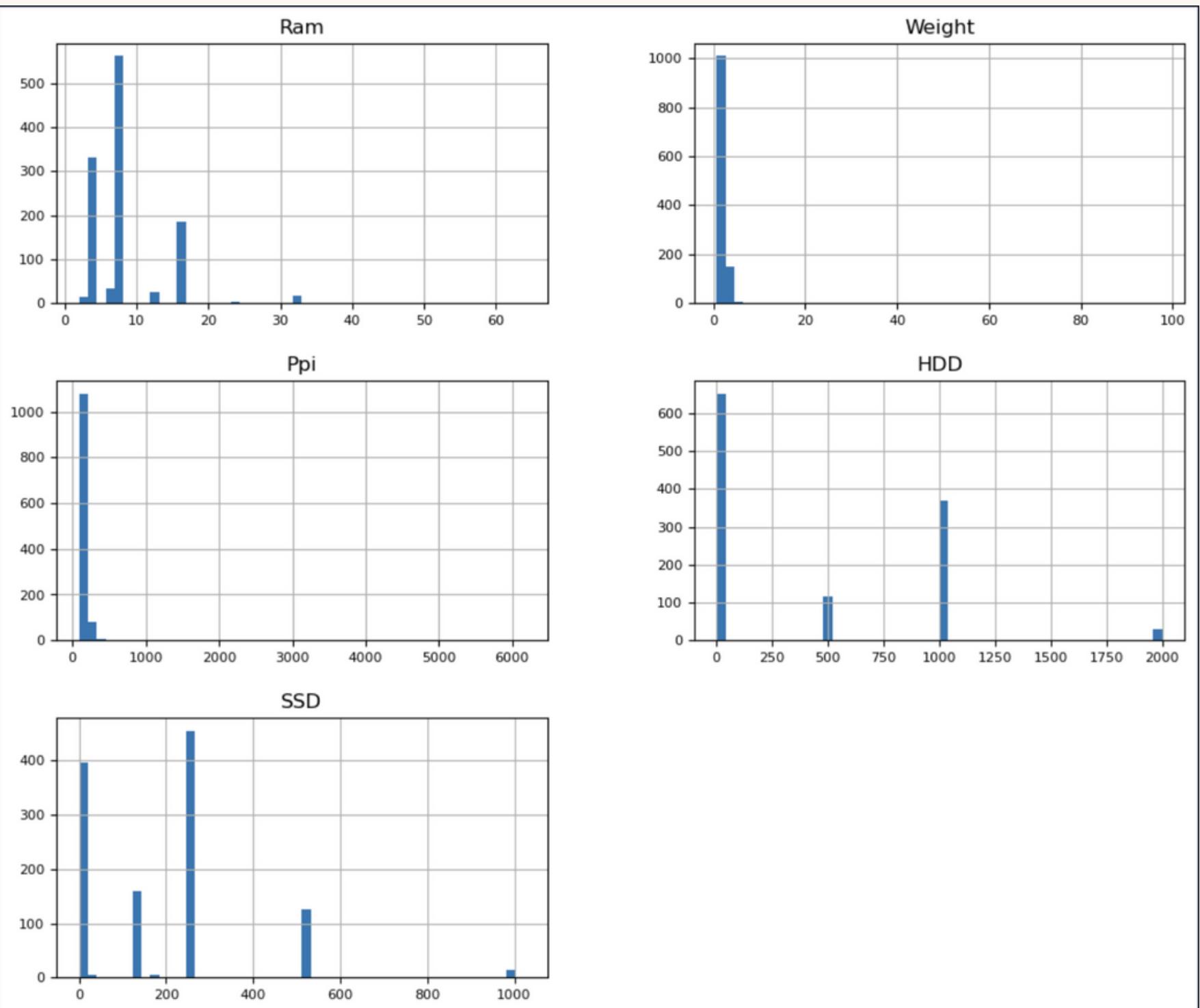
VALIDEZ:

Evaluamos el rango lógico y formsto correcto de valores en variables numericas, donde encontramos valores atípicos en algunas variables en lo cual utilizamos la media para reemplazar los valores a las variables atípicos

Búsqueda de relaciones entre las variables

En nuestro conjunto de datos, hemos observado una amplia variabilidad en los precios.

El almacenamiento SSD se relaciona positivamente con precios más altos. Aunque la cantidad de RAM no muestra una correlación lineal clara con los precios.



Interpretación de los coeficientes

Regresión lineal con outlier y sin normalizar

A partir de la interpretación de los coeficientes en el primero modelo de regresión con variables sin normalizar y con outliers, las variables que más afectan el precio son:

- **Type Name**
- **Almacenamiento SSD**
- **Company**
- **Touch Screen**
- **Cpu-brand**
- **Os_Windows**

Pero su intercepto es -1,539 lo que espera que el precio sea menor de 0 lo cual no es aceptable para este caso

	Coefficiente
Ram	-0.241888
Weight	-0.046949
Ppi	-0.000701
HDD	0.000076
SSD	0.016867
Company_Asus	0.829797
Company_Dell	1.085120
Company_HP	0.776862
Company_Lenovo	3.333099
Company_MSI	9.561444
Company_Mediacom	3.233462
Company_Others_Company	-3.696190
Company_Toshiba	-2.664050
Type Name_Gaming	14.656239
Type Name_Netbook	37.374408
Type Name_Notebook	14.458889
Type Name_Others_Type Name	13.503008
Type Name_Ultrabook	13.982586
Touch Screen_1	10.119085
Ips_1	1.886897
Cpu_brand_Intel Core i3	4.518321
Cpu_brand_Intel Core i5	2.758972
Cpu_brand_Intel Core i7	4.399733
Cpu_brand_Other Intel Processor	-1.564593
Gpu_brand_Intel	-1.168980
Gpu_brand_Nvidia	-4.491423
Os_Windows	-1.735186

Interpretación de los coeficientes

Regresión lineal sin outlier

Los coeficientes demuestran que las variables como una unidad SSD, la marca y modelo del procesador, el tipo de laptop y la presencia de características como la pantalla táctil y el sistema operativo Windows tienen un impacto significativo en el precio de una laptop.

Así mismo, el intercepto dio como resultado -0.00713 que representa como el precio base de un laptop si todas las variables llegaran a ser 0 pero comparado con el modelo anterior con outlier es un rango aceptable y es ideal para la predicción.

	Coefficiente
Ram	0.140790
SSD	0.191605
Company_Dell	0.050322
Company_HP	0.071650
Company_Toshiba	0.084739
TypeName_Netbook	-0.089906
TypeName_Notebook	-0.261756
TouchScreen_1	-0.065838
Cpu_brand_Intel Core i3	0.142259
Cpu_brand_Intel Core i5	0.541358
Cpu_brand_Intel Core i7	0.540408
Cpu_brand_Other Intel Processor	-0.095288
Os_Windows	0.113809

Métricas de calidad

MAE

Regresión lineal con outlier y sin normalizar

- Train: 5.6368
- Test: 4.4593

Regresión lineal sin outlier y con normalizar

- Train: 0.367
- Test: 0.449

RMSE

Regresión lineal con outlier y sin normalizar

- Train: 26.7476
- Test: 7.7408

Regresión lineal con outlier y sin normalizar

- Train: 0.469
- Test: 0.580

R²

Regresión lineal con outlier y sin normalizar

- Train: 0.04356
- Test: -156.4134

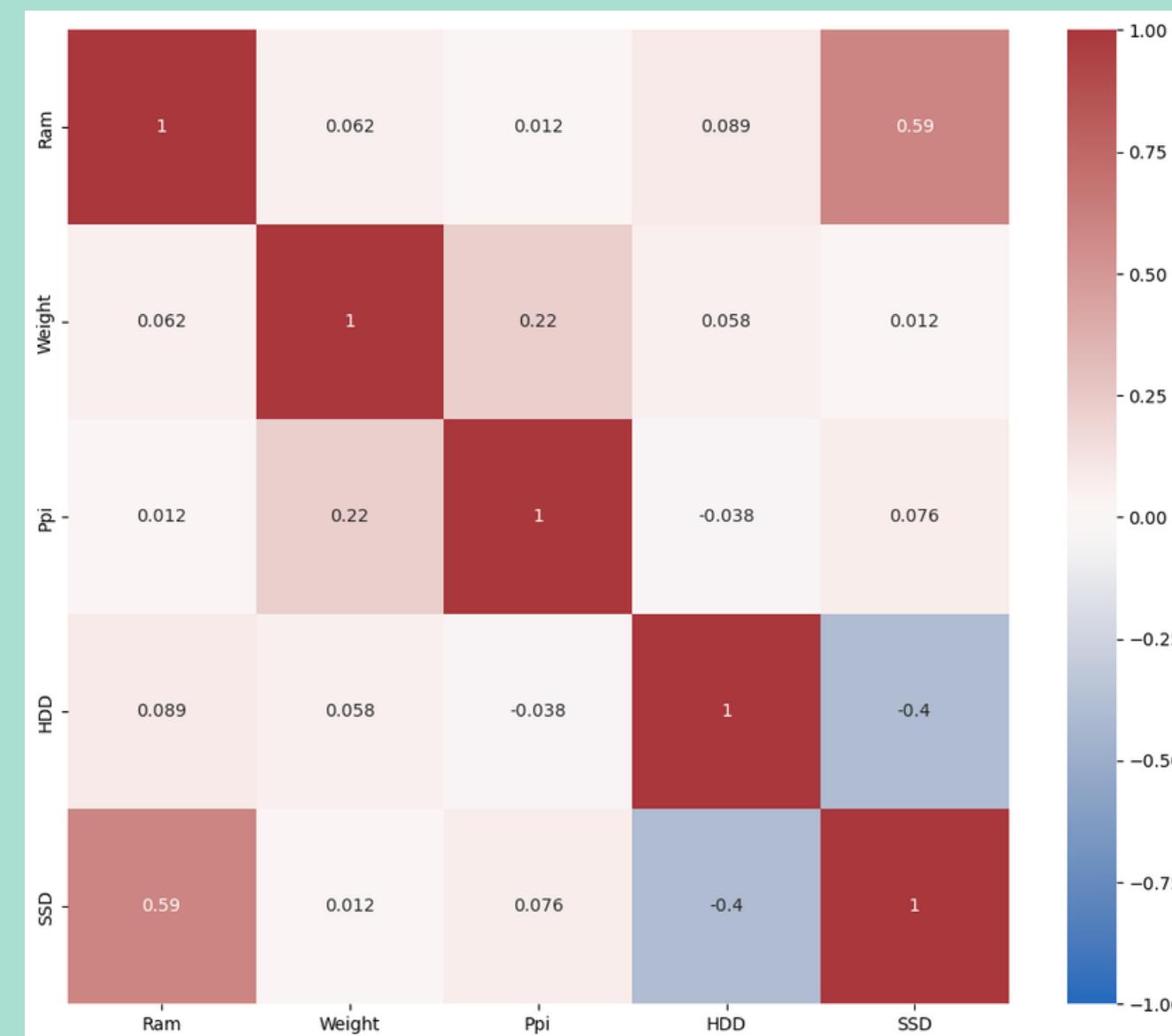
Regresión lineal sin outlier y con normalizar

- Train: 0.781
- Test: 0.650

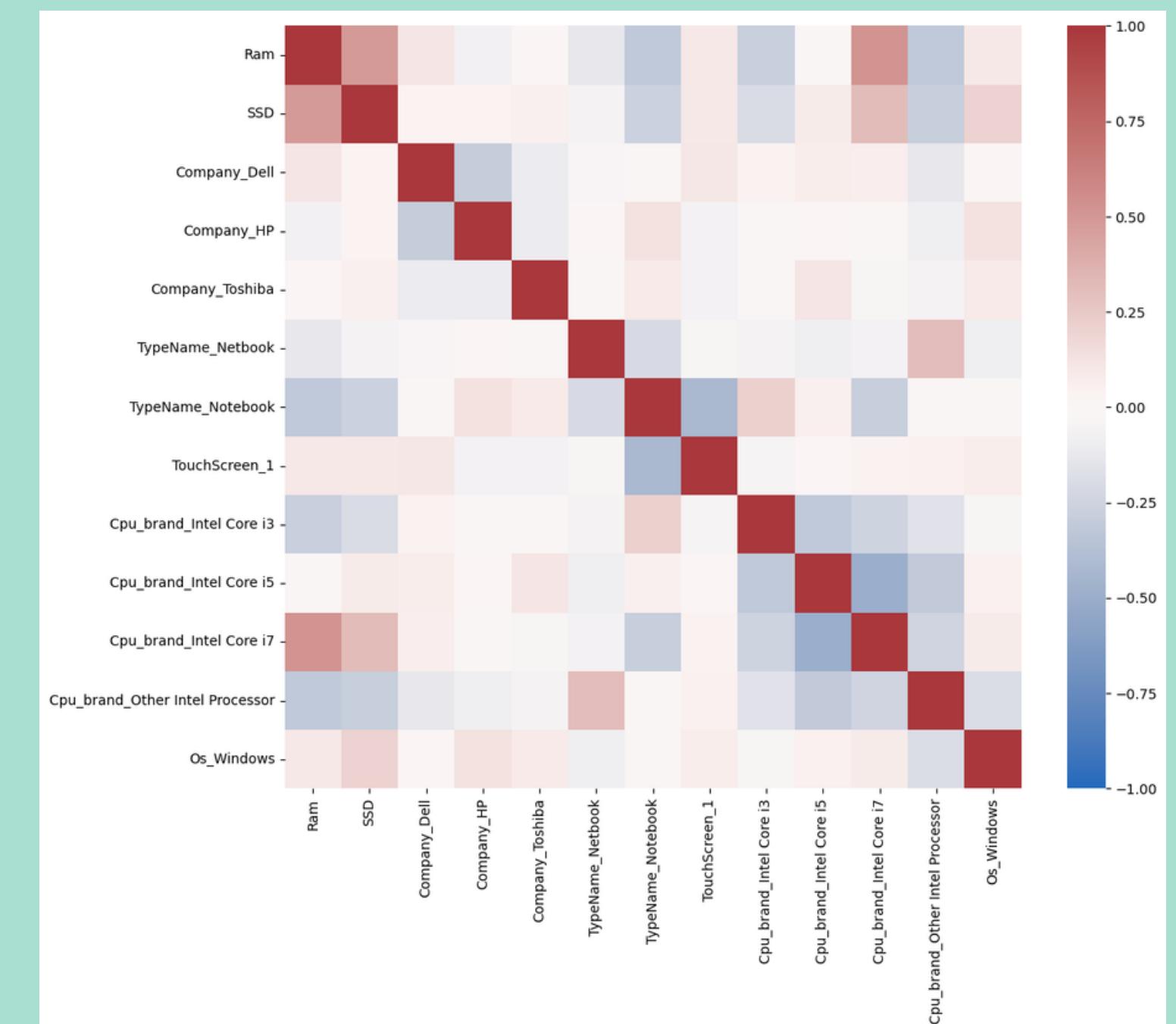
Comprobación de supuestos

Correlaciones entre las variables

Correlaciones del modelo con outliers

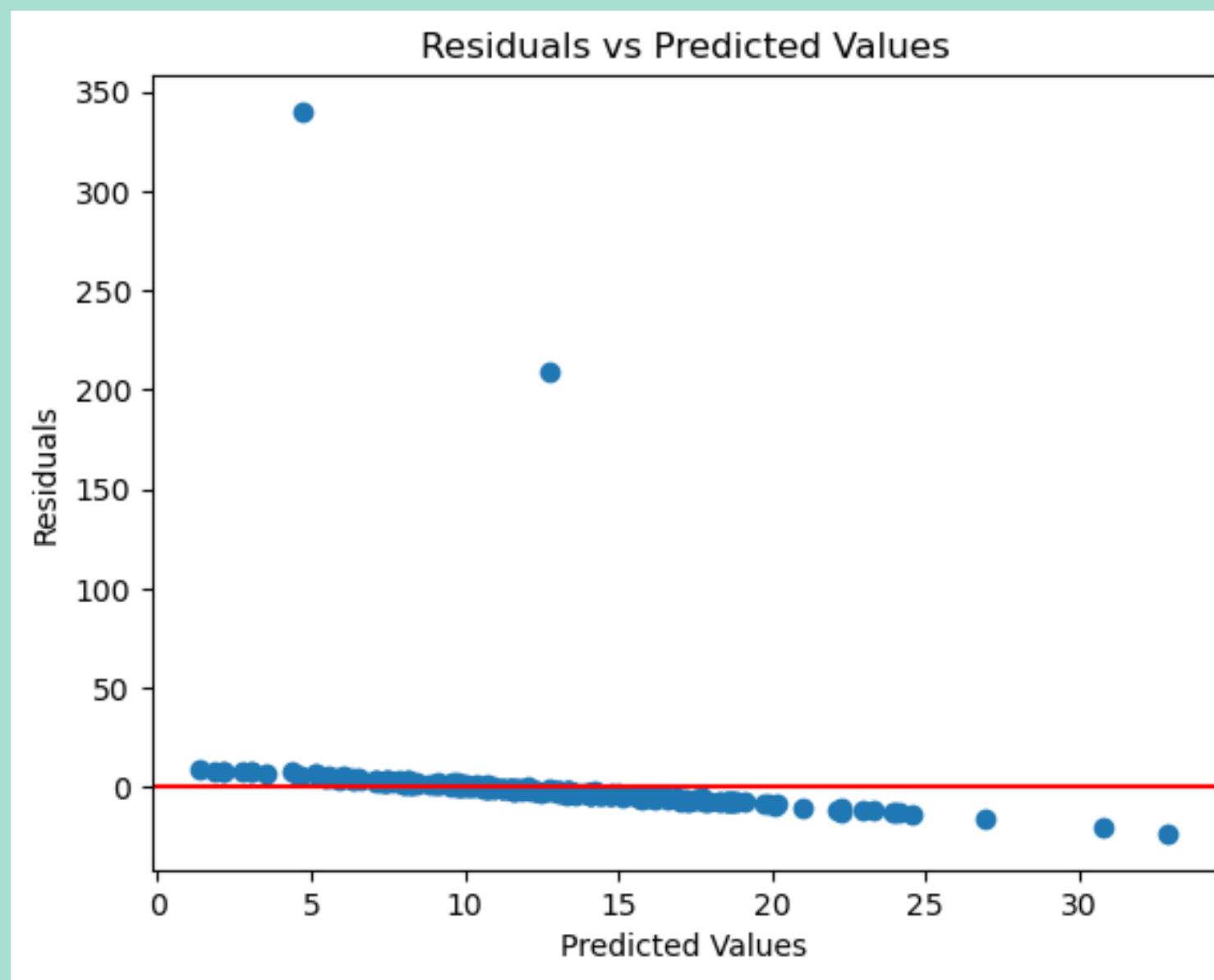


Correlaciones del modelo sin outliers

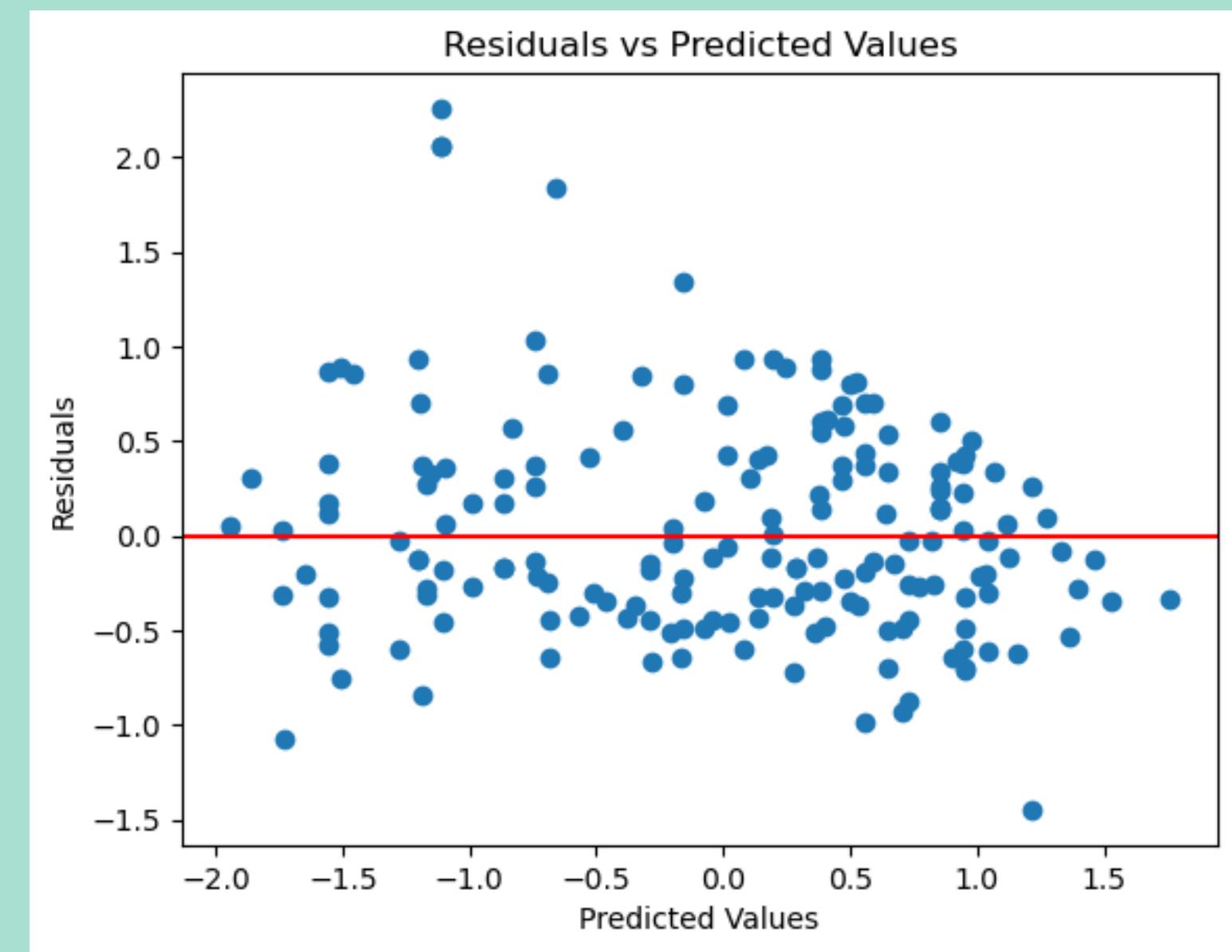


Homocedasticidad

Modelo con outliers

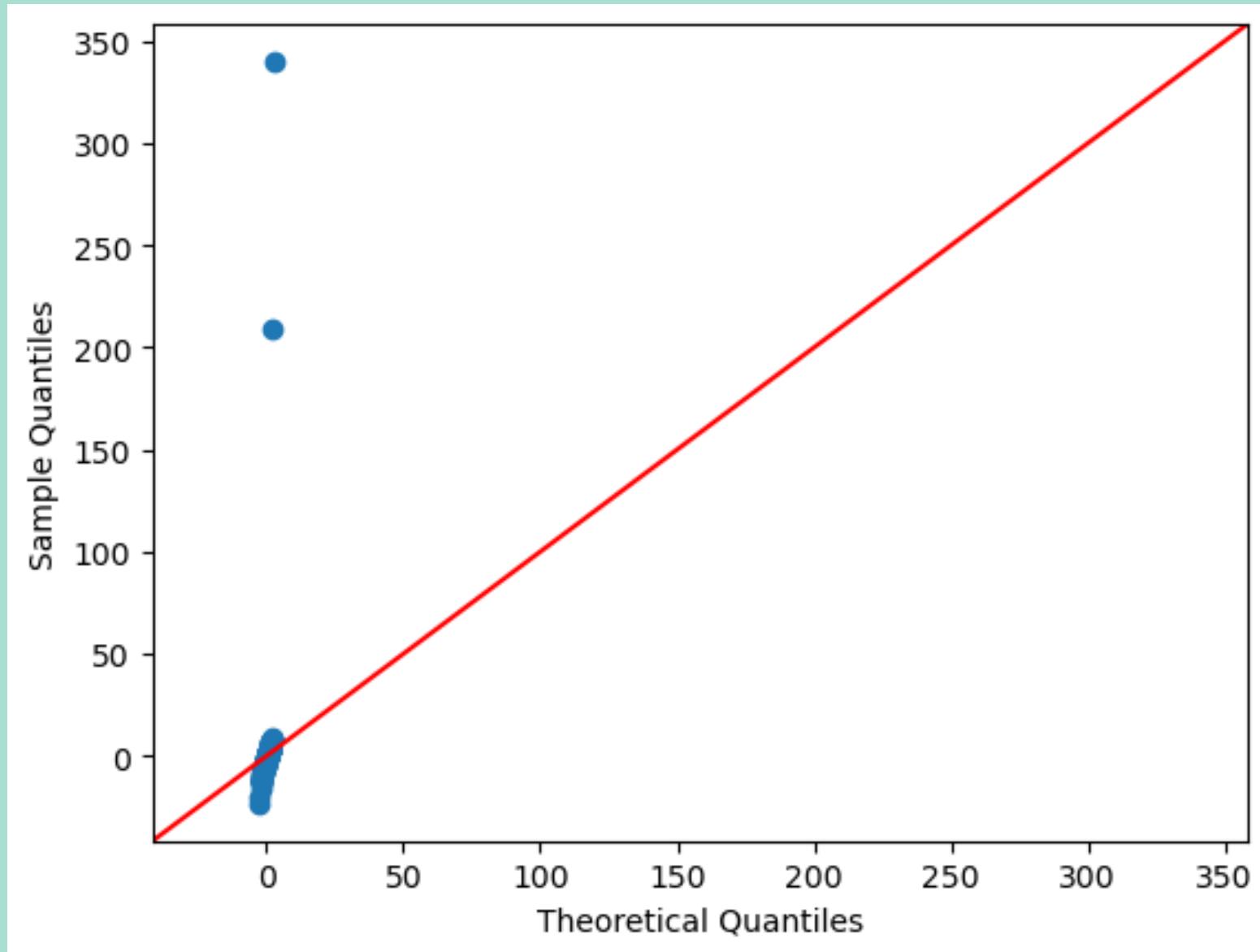


Modelo sin outliers

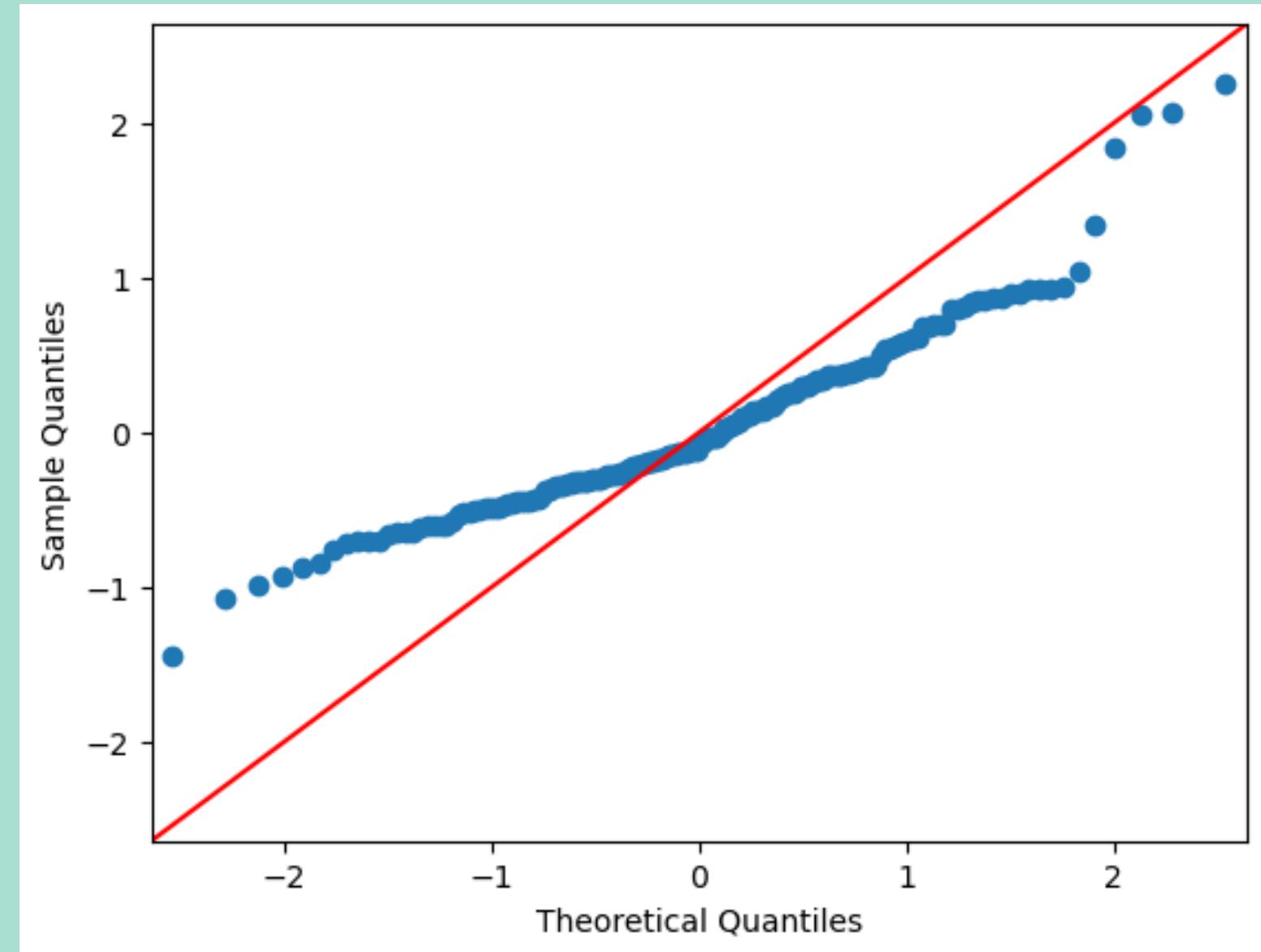


Normalidad

Modelo con outliers



Modelo sin outliers



Multicolinealidad

Modelo con outliers

	Variable	VIF
0	Ram	9.616862
1	Weight	1.359328
2	Ppi	1.306606
3	HDD	3.123224
4	SSD	5.350415
5	Company_Asus	2.425010
6	Company_Dell	4.357929
7	Company_HP	3.566636
8	Company_Lenovo	3.437270
9	Company_MSI	2.161575
10	Company_Mediacom	1.131281
11	Company_Others_Company	1.877394
12	Company_Toshiba	1.549608
13	TypeName_Gaming	6.437349
14	TypeName_Netbook	1.416267
15	TypeName_Notebook	10.969461
16	TypeName_Others_TypeName	1.690121
17	TypeName_Ultrabook	3.560464
18	TouchScreen_1	2.623343
19	Ips_1	1.880312
20	Cpu_brand_Intel Core i3	4.013623
21	Cpu_brand_Intel Core i5	10.340958
22	Cpu_brand_Intel Core i7	14.435187
23	Cpu_brand_Other Intel Processor	4.615628
24	Gpu_brand_Intel	9.647888
25	Gpu_brand_Nvidia	6.329623
26	Os_Windows	8.369853

Modelo sin outliers

	Variable	VIF
0	Ram	1.868218
1	SSD	1.444278
2	Company_Dell	1.210166
3	Company_HP	1.208251
4	Company_Toshiba	1.077655
5	TypeName_Netbook	1.230349
6	TypeName_Notebook	1.644090
7	TouchScreen_1	1.316231
8	Cpu_brand_Intel Core i3	3.056453
9	Cpu_brand_Intel Core i5	4.934716
10	Cpu_brand_Intel Core i7	4.841320
11	Cpu_brand_Other Intel Processor	3.162002
12	Os_Windows	1.114180

Modelo Predictivo

Aplicamos el modelo final con los datos objetivos con el fin de hallar el precio de laptops, con la evaluación del modelo y con los errores predictivos. El modelo de predicción del precio de laptops sin outliers y con variables normalizadas destaca como la opción más sólida y precisa. La eliminación de valores atípicos permite al modelo enfocarse en patrones de datos representativos, evitando distorsiones en las predicciones. La normalización de variables garantiza que todas las características se consideren equitativamente, facilitando la interpretación de cómo cada una influye en el precio. En contraste, los modelos que no tratan los outliers o no normalizan las variables pueden verse afectados por valores extremos y dificultades en la interpretación de los coeficientes.

