

1) GLM & Stepwise both con Precio original

formula <- price ~ floor + size + rooms + bathrooms + latitude + longitude + hasLift +

distancia_minima_banco + distancia_minima_starbucks + distancia_minima_atractivos +

distancia_minima_buenasmigas + distancia_minima_clubesnocturnos +

distancia_minima_mcdonalds + distancia_minima_metro + distancia_minima_parking +

distancia_minima_parkingbici + distancia_minima_parques + distancia_minima_playa +

distancia_minima_santagloria + distancia_minima_vivari + num_airbnbs_500 +

media_precios_airbnbs_500

Entrenamiento:

Suma de los errores al cuadrado (SSE): 18796119037

R-squared: 0.2516718

AIC: 189371.4

Prueba:

R-squared ajustado para el conjunto de prueba: 0.08234617

Resultado Stepwise both:

```
call:
glm(formula = price ~ floor + size + bathrooms + latitude + hasLift +
  distancia_minima_banco + distancia_minima_starbucks + distancia_minima_atractivos +
  distancia_minima_buenasmigas + distancia_minima_clubesnocturnos +
  distancia_minima_mcdonalds + distancia_minima_metro + distancia_minima_parkingbici +
  distancia_minima_parques + distancia_minima_playa + distancia_minima_santagloria +
  distancia_minima_vivari, family = gaussian, data = train_data)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.403e+04	1.302e+04	4.919	8.83e-07 ***
floor	2.278e+01	1.060e+01	2.149	0.031622 *
size	1.413e+01	1.023e+00	13.806	< 2e-16 ***
bathrooms	1.605e+03	6.158e+01	26.059	< 2e-16 ***
latitude	-1.493e+03	3.144e+02	-4.748	2.08e-06 ***
hasLift	2.891e+02	2.575e+01	11.229	< 2e-16 ***
distancia_minima_banco	-3.041e-01	7.491e-02	-4.059	4.95e-05 ***
distancia_minima_starbucks	-5.291e-02	3.470e-02	-1.525	0.127402
distancia_minima_atractivos	-2.387e-01	6.403e-02	-3.728	0.000194 ***
distancia_minima_buenasmigas	6.895e-01	4.771e-02	14.453	< 2e-16 ***
distancia_minima_clubesnocturnos	4.843e-01	4.663e-02	10.384	< 2e-16 ***
distancia_minima_mcdonalds	-3.817e-01	5.119e-02	-7.457	9.55e-14 ***
distancia_minima_metro	2.010e-01	8.543e-02	2.353	0.018635 *
distancia_minima_parkingbici	5.486e-01	9.530e-02	5.757	8.81e-09 ***
distancia_minima_parques	1.668e-01	8.822e-02	1.891	0.058716 .
distancia_minima_playa	-3.678e-01	2.027e-02	-18.148	< 2e-16 ***
distancia_minima_santagloria	-1.158e+00	3.508e-02	-33.000	< 2e-16 ***
distancia_minima_vivari	4.156e-01	3.671e-02	11.322	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 1709560)

Null deviance: 2.5117e+10 on 11014 degrees of freedom

Residual deviance: 1.8800e+10 on 10997 degrees of freedom

AIC: 189364

Number of Fisher Scoring iterations: 2

Suma de los errores al cuadrado (SSE) del modelo stepwise: 6667732530

R-squared del modelo stepwise: 0.08860653

AIC del modelo stepwise: 189363.7

2) GLM & Stepwise both con Precio logaritmico

formula <- price ~ floor + size + rooms + bathrooms + latitude + longitude + hasLift +

distancia_minima_banco + distancia_minima_starbucks + distancia_minima_atractivos +

distancia_minima_buenasmigas + distancia_minima_clubesnocturnos +

distancia_minima_mcdonalds + distancia_minima_metro + distancia_minima_parking +

distancia_minima_parkingbici + distancia_minima_parques + distancia_minima_playa +

distancia_minima_santagloria + distancia_minima_vivari + num_airbnbs_500 +

media_precios_airbnbs_500

Entrenamiento:

Suma de los errores al cuadrado (SSE): 2381.83

R-squared: 0.3432204

AIC: 14438.97

Prueba:

R-squared ajustado para el conjunto de prueba: 0.08887699

Resultado Stepwise both:

```
call:
glm(formula = price ~ floor + size + bathrooms + latitude + hasLift +
  distancia_minima_banco + distancia_minima_atractivos + distancia_minima_buenasmigas +
  distancia_minima_clubesnocturnos + distancia_minima_mcdonalds +
  distancia_minima_metro + distancia_minima_parkingbici + distancia_minima_parques +
  distancia_minima_playa + distancia_minima_santagloria + distancia_minima_vivari,
  family = gaussian, data = train_data)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	5.231e+01	4.633e+00	11.291	< 2e-16 ***
floor	1.076e-02	3.773e-03	2.851	0.00437 **
size	6.435e-03	3.643e-04	17.665	< 2e-16 ***
bathrooms	5.688e-01	2.192e-02	25.946	< 2e-16 ***
latitude	-1.081e+00	1.119e-01	-9.655	< 2e-16 ***
hasLift	1.240e-01	9.159e-03	13.541	< 2e-16 ***
distancia_minima_banco	-2.098e-04	2.665e-05	-7.871	3.84e-15 ***
distancia_minima_atractivos	-1.361e-04	2.279e-05	-5.973	2.40e-09 ***
distancia_minima_buenasmigas	2.157e-04	1.696e-05	12.715	< 2e-16 ***
distancia_minima_clubesnocturnos	1.926e-04	1.660e-05	11.600	< 2e-16 ***
distancia_minima_mcdonalds	-1.690e-04	1.797e-05	-9.404	< 2e-16 ***
distancia_minima_metro	1.066e-04	3.041e-05	3.504	0.00046 **
distancia_minima_parkingbici	2.829e-04	3.337e-05	8.476	< 2e-16 ***
distancia_minima_parques	1.370e-04	3.121e-05	4.391	1.14e-05 **
distancia_minima_playa	-1.321e-04	6.674e-06	-19.800	< 2e-16 ***
distancia_minima_santagloria	-4.692e-04	1.195e-05	-39.259	< 2e-16 ***
distancia_minima_vivari	1.821e-04	1.238e-05	14.716	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for gaussian family taken to be 0.2166645)

Null deviance: 3626.5 on 11014 degrees of freedom
Residual deviance: 2382.9 on 10998 degrees of freedom
AIC: 14432

Number of Fisher Scoring iterations: 2

Suma de los errores al cuadrado (SSE) del modelo stepwise: 759.9733

R-squared del modelo stepwise: 0.09613749

AIC del modelo stepwise: 14431.81

***Como estamos calculando el log(price) es necesario convertir price a la escala original para calcular las métricas:**

Suma de los errores al cuadrado (SSE) del modelo stepwise: 759.9733

Mean Squared Error (MSE) del modelo stepwise: 0.248601

3) GLM & Stepwise both con log(price) en formula

```
log(price) ~ floor + size + rooms + bathrooms + latitude + longitude + hasLift +  
distancia_minima_banco + distancia_minima_starbucks + distancia_minima_atractivos +  
distancia_minima_buenasmigas + distancia_minima_clubesnocturnos +  
distancia_minima_mcdonalds + distancia_minima_metro + distancia_minima_parking +  
distancia_minima_parkingbici + distancia_minima_parques + distancia_minima_playa +  
distancia_minima_santagloria + distancia_minima_vivari + num_airbnbs_500 +  
media_precios_airbnbs_500
```

Entrenamiento:

Suma de los errores al cuadrado (SSE): 96391181578

R-squared: -2.837614

AIC: 14438.97

Prueba:

R-squared ajustado para el conjunto de prueba: -2.834001

Resultado Stepwise both:

```
Call:  
glm(formula = log(price) ~ floor + size + bathrooms + latitude +  
hasLift + distancia_minima_banco + distancia_minima_atractivos +  
distancia_minima_buenasmigas + distancia_minima_clubesnocturnos +  
distancia_minima_mcdonalds + distancia_minima_metro + distancia_minima_parkingbici +  
distancia_minima_parques + distancia_minima_playa + distancia_minima_santagloria +  
distancia_minima_vivari, family = gaussian, data = train_data)  
  
Coefficients:  
              Estimate Std. Error t value Pr(>|t|)  
(Intercept)    5.231e+01  4.633e+00  11.291 < 2e-16 ***  
floor           1.076e-02  3.773e-03   2.851  0.00437 **  
size            6.435e-03  3.643e-04  17.665 < 2e-16 ***  
bathrooms       5.688e-01  2.192e-02  25.946 < 2e-16 ***  
latitude        -1.081e+00  1.119e-01  -9.655 < 2e-16 ***  
haslift         1.240e-01  9.159e-03  13.541 < 2e-16 ***  
distancia_minima_banco -2.098e-04  2.665e-05  -7.871 3.84e-15 ***  
distancia_minima_atractivos -1.361e-04  2.279e-05  -5.973 2.40e-09 ***  
distancia_minima_buenasmigas 2.157e-04  1.696e-05  12.715 < 2e-16 ***  
distancia_minima_clubesnocturnos 1.926e-04  1.660e-05  11.600 < 2e-16 ***  
distancia_minima_mcdonalds -1.690e-04  1.797e-05  -9.404 < 2e-16 ***  
distancia_minima_metro 1.066e-04  3.041e-05   3.504  0.00046 ***  
distancia_minima_parkingbici 2.829e-04  3.337e-05   8.476 < 2e-16 ***  
distancia_minima_parques 1.370e-04  3.121e-05   4.391 1.14e-05 ***  
distancia_minima_playa -1.321e-04  6.674e-06 -19.800 < 2e-16 ***  
distancia_minima_santagloria -4.692e-04  1.195e-05 -39.259 < 2e-16 ***  
distancia_minima_vivari 1.821e-04  1.238e-05  14.716 < 2e-16 ***  
---  
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1  
  
(Dispersion parameter for gaussian family taken to be 0.2166645)  
  
Null deviance: 3626.5  on 11014  degrees of freedom  
Residual deviance: 2382.9  on 10998  degrees of freedom  
AIC: 14432  
  
Number of Fisher Scoring iterations: 2
```

Suma de los errores al cuadrado (SSE) del modelo stepwise: 27847531502

R-squared del modelo stepwise: -2.8064

AIC del modelo stepwise: 14431.81

***Como estamos calculando el log(price) es necesario convertir price a la escala original para calcular las métricas:**

Suma de los errores al cuadrado (SSE) del modelo stepwise: 27847531502

Mean Squared Error (MSE) del modelo stepwise: 9109431

Resumen

		Precio original	Precio en base log	log(price)
Entrenamiento	SSE	18796119037	2382	96391181578
	R-cuadrado	0,25	0,34	-2,84
	AIC	189371	14439	14439
Prueba	SSE	6665202735	760563	27847533095
	R-cuadrado	0,082	0,095	-2,81
	AIC	189371	14439	14438.97
Stepwise both	SSE Stepwise	6667732530	7599733	27847531502
	R-cuadrado Stepwise	0,09	0,10	-2,81
	AIC stepwise	189364	14432	14432
Stepwise both escala original	SSE Stepwise		7.599.733	27847531502
	R-cuadrado Stepwise		0,10	-2,81
	AIC stepwise		14432	14432