

Predicción de Asistencia a Festivales en Argentina

Aplicación de técnicas de regresión supervisada sobre datos del SInCA

Tecnicatura Superior en Ciencia de Datos e IA
Estudiante: Mariana Anahí López

Año 2025



CENTRO POLITÉCNICO SUPERIOR
MALVINAS ARGENTINAS

- Los festivales y fiestas populares son parte esencial de la identidad argentina.
- Generan impacto turístico, económico y social en todo el país.
- No existe información precisa sobre la cantidad de asistentes a cada evento.
- Predecirla permite mejorar la planificación, promoción y gestión.





Objetivo general:

Predecir la cantidad aproximada de asistentes a festivales en Argentina a partir de sus características descriptivas.

Objetivos específicos:

- Analizar el dataset del SInCA.
- Limpiar y codificar variables categóricas.
- Entrenar y comparar modelos de regresión.
- Evaluar desempeño mediante métricas.
- Aplicar el modelo final a casos sin datos.



- Fuente: Catálogo Nacional de Fiestas y Festivales – SInCA
- Registros: 3255 eventos en todo el país
- Variables principales:
- Provincia
- Temática principal
- Tipo de gestión
- Tipo de entrada
- Duración (días)
- Mes de realización
- Cantidad aproximada de asistentes (variable objetivo)
- ~50% de valores faltantes en la variable objetivo





Origen y características de los datos

1	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Columna1	cod_dep	cod_loc	Categoría	Provincia	Departamento	Localidad/Paraje	Nombre	Latitud	Longitud	Nota	Fuente	Temática_principa	Temática_secundaria	Aniversario
2	6	6778	6778020	Fiesta y/o fest	Buenos Aires	San Vicente	San Vicente	Día Nacional	-3502403944	-5842059144	Centroide	Festivales Arg	Tradicionales	s/d	No
3	6	6497	6497060	Fiesta y/o fest	Buenos Aires	Lujan	Lujan	Festival del Pa	-3457065507	-5910954018	Centroide	Festivales Arg	Gastronómica	s/d	No
4	6	6196	6196010	Fiesta y/o fest	Buenos Aires	Coronel Pringles	Coronel Pringles	Pringles Cine I	-3798652106	-6135407157	Centroide	Festivales Arg	Artística	s/d	No
5	6	6280	6280040	Fiesta y/o fest	Buenos Aires	General Alvarado	Miramar	Fiesta del Kiwi	-3827074293	-5784047146	Centroide	RECREO - Got	Productiva	s/d	No
6	6	6112	6112010	Fiesta y/o fest	Buenos Aires	Bragado	Bragado	Suena Rock	-3511894230	-6048791476	Centroide	Festivales Arg	Artística	s/d	No
7	6	6441	6441030	Carnaval	Buenos Aires	La Plata	La Plata	Carnaval Disic	-3492206666	-5795439165	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
8	6	6351	6351020	Fiesta y/o fest	Buenos Aires	General Pinto	General Pinto	FESTIVAL DE II	-3476415599	-6189006749	Centroide	Festivales Arg	Artística	s/d	No
9	6	6532	6532020	Fiesta y/o fest	Buenos Aires	Mercedes	Mercedes	Festival de Tar	-3465217154	-5942959817	Centroide	Festivales Arg	Artística	s/d	No
10	6	6490	6490010	Fiesta y/o fest	Buenos Aires	Lomas de Zamora	Lomas de Zamora	Festival Cultur	-3476116024	-5839876606	Centroide	Festivales Arg	Artística	s/d	No
11	6	6560	6560010	Fiesta y/o fest	Buenos Aires	Moreno	Moreno	FESTIVAL URB	-3464845620	-5878703763	Centroide	Festivales Arg	Artística	s/d	No
12	6	6763	6763050	Fiesta y/o fest	Buenos Aires	San Nicolas	San Nicolas de los A	FESTIVAL RIC	-3333469892	-6021839745	Centroide	Festivales Arg	Gastronómica	s/d	No
13	6	6490	6490010	Fiesta y/o fest	Buenos Aires	Lomas de Zamora	Lomas de Zamora	FESTIVAL LATI	-3476116024	-5839876606	Centroide	Festivales Arg	s/d	s/d	No
14	6	6658	6658010	Fiesta y/o fest	Buenos Aires	Quilmes	Quilmes	FESTIVAL ENT	-3472469143	-5826134074	Centroide	Festivales Arg	s/d	s/d	No
15	6	6553	6553020	Carnaval	Buenos Aires	Moreno	Moreno	Carnavales 20	-3898150654	-6130054575	Centroide	Festivales Arg	Artística	s/d	No
16	6	6441	6441030	Carnaval	Buenos Aires	La Plata	Lomas de Copello	Fiesta de Carr	-3492206666	-5795439165	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
17	6	6861	6861010	Fiesta y/o fest	Buenos Aires	Vicente Lopez	Vicente Lopez	FESTIVAL DE II	-3450850085	-5848163942	Centroide	Festivales Arg	Artística	s/d	No
18	6	6469	6469080	Fiesta y/o fest	Buenos Aires	Lincoln	Lincoln	Encuentro de l	-3486904222	-6152916497	Centroide	RECREO - Got	Artística	s/d	No
19	6	6245	6245010	Fiesta y/o fest	Buenos Aires	Ensenada	Ensenada	Festival Intern	-3485908765	-5791320630	Centroide	Festivales Arg	Artística	s/d	No
20	6	6441	6441030	Carnaval	Buenos Aires	La Plata	La Plata	Carnaval del C	-3492206666	-5795439165	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
21	6	6638	6638040	Carnaval	Buenos Aires	Pilar	Pilar	Carnaval Mun	-3445648210	-5891472419	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
22	6	6357	6357110	Fiesta y/o fest	Buenos Aires	General Pueyrred	Mar del Plata	¡QUE ARDA! F	-3799886401	-5754753294	Centroide	Festivales Arg	Otra	Artística	No
23	6	6497	6497060	Carnaval	Buenos Aires	Lujan	Lujan	Carnaval en Lu	-3457065507	-5910954018	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
24	6	6791	6791050	Carnaval	Buenos Aires	Tandil	Tandil	Carnaval Barri	-3732388491	-5913106918	Centroide	Carnavales Ar	Expresiones del carna	s/d	No
25	6	6539	6539010	Fiesta y/o fest	Buenos Aires	Merlo	Merlo	Festival Intern	-3466733329	-5872849022	Centroide	Festivales Arg	Artística	s/d	No
26	6	6490	6490010	Fiesta y/o fest	Buenos Aires	Lomas de Zamora	Lomas de Zamora	Festival Alborc	-3476116024	-5839876606	Centroide	Festivales Arg	Artística	s/d	No
27	6	6518	6518040	Fiesta y/o fest	Buenos Aires	Mar Chiquita	Mar Chiquita	Festival de Ave	-3774625817	-5742744793	Centroide	Festivales Arg	Otra	s/d	No
28	6	6644	6644010	Fiesta y/o fest	Buenos Aires	Pinamar	Pinamar	FLIP FESTIVAL	-3710354366	-5684842825	Centroide	Festivales Arg	Artística	s/d	No
29	6	6497	6497060	Fiesta y/o fest	Buenos Aires	Lujan	Lujan	FestiAcal	-3457065507	-5910954018	Centroide	Festivales Arg	Artística	s/d	No
30	6	6868	6868020	Fiesta y/o fest	Buenos Aires	Villa Gesell	Villa Gesell	6ta Fiesta Trac	-3724647047	-5697072676	Centroide	RECREO - Got	Tradicionales	s/d	No
31	6	6063	6063010	Fiesta y/o fest	Buenos Aires	Balcarce	Balcarce	Fiesta de la Pr	-3784827793	-5825516658	Centroide	Festivales Arg	Artística	s/d	No
32	6	6021	6021010	Fiesta y/o fest	Buenos Aires	Alberti	Alberti	Fiesta Provinc	-3503307343	-6028061973	Centroide	Festivales Arg	Artística	s/d	No
33	6	6595	6595070	Carnaval	Buenos Aires	Olavarria	Olavarria	Carnaval Abi	-3699200259	-6031800466	Centroide	Carnavales Ar	Expresiones del carna	s/d	No



Origen y características de los datos

	Q	R	S	T	U	V	W	X	Y	Z	AA
1	Tipos de gestión	privada	Periodicidad	Año primera edición	Ediciones	Mes de realización	Semana de realización	Duración (días)	Cantidad aprox de asistentes	Tipo Entrada	Modalidad
2	s/d	s/d	s/d	0	Julio	4	s/d	1000	Gratuita	s/d	2022
3	s/d	s/d	s/d	1	Diciembre	3	s/d	2000	Gratuita	s/d	2022
4	s/d	s/d	s/d	1	Noviembre	2	s/d	250	Gratuita	s/d	2022
5	s/d	s/d	s/d	1	s/d	s/d	s/d	s/d	s/d	s/d	2022
6	s/d	s/d	s/d	2	Febrero	3	s/d	3000	Gratuita	s/d	2023
7	s/d	s/d	s/d	2	Febrero	3	s/d	300	Gratuita	s/d	2023
8	s/d	s/d	s/d	2	Julio	3	s/d	6000	Gratuita	s/d	2022
9	s/d	s/d	s/d	2	Marzo	3	s/d	7000	Gratuita	s/d	2023
10	s/d	s/d	2008	2	Noviembre	4	s/d	529	Paga	s/d	2022
11	s/d	s/d	s/d	2	Octubre	1	s/d	50000	Gratuita	s/d	2022
12	s/d	s/d	s/d	2	Octubre	1	s/d	45000	Gratuita	s/d	2022
13	Sin fines de lucro	s/d	s/d	2	s/d	s/d	s/d	s/d	s/d	s/d	2021
14	Sin fines de lucro	s/d	s/d	2	s/d	s/d	s/d	s/d	s/d	s/d	2021
15	s/d	s/d	s/d	3	Enero	1	s/d	60000	Gratuita	s/d	2023
16	s/d	s/d	s/d	3	Febrero	3	s/d	NA	Gratuita	s/d	2023
17	s/d	s/d	2019	3	Noviembre	4	s/d	400	Gratuita	s/d	2022
18	s/d	s/d	s/d	3	s/d	s/d	s/d	s/d	s/d	s/d	2022
19	s/d	s/d	s/d	3	Septiembre	3	s/d	1500	Gratuita	s/d	2022
20	s/d	s/d	s/d	4	Febrero	3	s/d	500	Gratuita	s/d	2023
21	s/d	s/d	s/d	4	Febrero	1	s/d	NA	Gratuita	s/d	2023
22	s/d	s/d	s/d	4	Marzo	1	s/d	300	Gratuita	s/d	2023
23	s/d	s/d	s/d	5	Febrero	2	s/d	NA	Gratuita	s/d	2023
24	s/d	s/d	s/d	5	Marzo	1	s/d	NA	Gratuita	s/d	2023
25	s/d	s/d	s/d	5	Mayo	2	s/d	6000	Gratuita	s/d	2023
26	s/d	s/d	2003	5	Noviembre	3	s/d	1000	Gratuita	s/d	2022
27	Sin fines de lucro	s/d	2017	5	Noviembre	1	s/d	2566	Gratuita	s/d	2022
28	s/d	s/d	s/d	6	Agosto	3	s/d	500	Gratuita	s/d	2022
29	s/d	s/d	s/d	6	Febrero	3	s/d	400	Paga	s/d	2023
30	s/d	s/d	s/d	6	s/d	s/d	s/d	s/d	s/d	s/d	2022
31	s/d	s/d	s/d	6	Septiembre	3	s/d	2500	Gratuita	s/d	2022
32	s/d	s/d	s/d	8	Diciembre	1	s/d	8000	Gratuita	s/d	2022
33	s/d	s/d	s/d	9	Marzo	1	s/d	2000	Gratuita	s/d	2022



Preparación de los datos

- Limpieza de duplicados y normalización de texto.
- Imputación de valores faltantes.
- Codificación de variables categóricas con One-Hot Encoding.
- Separación en conjuntos de entrenamiento y prueba.

Preparación de los datos



CENTRO POLÍTÉCNICO SUPERIOR
MALVINAS ARGENTINAS

Autoguardado H DATA SET CODIFICADO.festivales_codificado_mo... • Se guardó en Esta PC

Buscar

Archivo Inicio Insertar Trazo Disposición de página Fórmulas Datos Revisar Vista Automatizar Programador Ayuda Diseño de tabla Consulta

Comentarios Compartir

action_No	tipo_de_gestion	tipo_de_gestion_Público	tipo_de_gestion_Enero	mes_de_realizaci_n_Febrero	mes_de_realizaci_n_Marzo	mes_de_realizaci_n_Abril	mes_de_realizaci_n_Mayo	mes_de_realizaci_n_Junio	mes_de_realizaci_n_Julio	mes_de_realizaci_n_Agosto	mes_de_realizaci_n_Septiembre	mes_de_realizaci_n_Octubre	mes_de_realizaci_n_Noviembre	mes_de_realizaci_n_Diciembre
1	Privado													
2	0	0	1	0	0	0	0	0	0	1	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	0	0	0	1
5	1	0	0	0	0	0	0	0	0	0	0	0	0	1
6	1	0	0	0	1	0	0	0	0	0	0	0	0	0
7	0	1	0	0	1	0	0	0	0	0	0	0	0	0
8	0	0	1	0	0	0	0	0	0	1	0	0	0	0
9	0	0	1	0	0	1	0	0	0	0	0	0	0	0
10	0	1	0	0	0	0	0	0	0	0	0	0	0	1
11	0	0	1	0	0	0	0	0	0	0	0	0	0	1
12	0	0	1	0	0	0	0	0	0	0	0	0	1	0
13	0	1	0	0	0	0	0	0	0	0	0	0	0	1
14	0	1	0	0	0	0	0	0	0	0	0	0	0	1
15	1	0	0	1	0	0	0	0	0	0	0	0	0	0
16	0	1	0	0	1	0	0	0	0	0	0	0	0	0
17	0	1	0	0	0	0	0	0	0	0	0	0	0	1
18	1	0	0	0	0	0	0	0	0	0	0	0	0	1
19	0	0	1	0	0	0	0	0	0	0	0	1	0	0
20	0	1	0	0	1	0	0	0	0	0	0	0	0	0
21	0	1	0	0	1	0	0	0	0	0	0	0	0	0
22	1	0	0	0	0	1	0	0	0	0	0	0	0	0
23	0	1	0	0	1	0	0	0	0	0	0	0	0	0
24	0	1	0	0	0	1	0	0	0	0	0	0	0	0
25	1	0	0	0	0	0	0	1	0	0	0	0	0	0
26	0	1	0	0	0	0	0	0	0	0	0	0	0	1
27	0	1	0	0	0	0	0	0	0	0	0	0	0	1
28	0	0	1	0	0	0	0	0	0	0	1	0	0	0
29	1	0	0	0	1	0	0	0	0	0	0	0	0	0
30	1	0	0	0	0	0	0	0	0	0	0	0	0	1
31	0	0	1	0	0	0	0	0	0	0	0	1	0	0

DATA SET CODIFICADO_festivales_

DATA SET CODIFICADO.festivales_

+

Lista



Accesibilidad: No disponible

Promedio: 0.344393241 Recuento: 3255 Suma: 1121

100%

Modelos implementados

Modelo 1 – Regresión Lineal Múltiple:

- Base de comparación.
- Interpretable, útil para analizar el peso de variables.

Modelo 2 – Random Forest Regressor:

- Capta relaciones no lineales.
- Maneja interacciones entre variables.

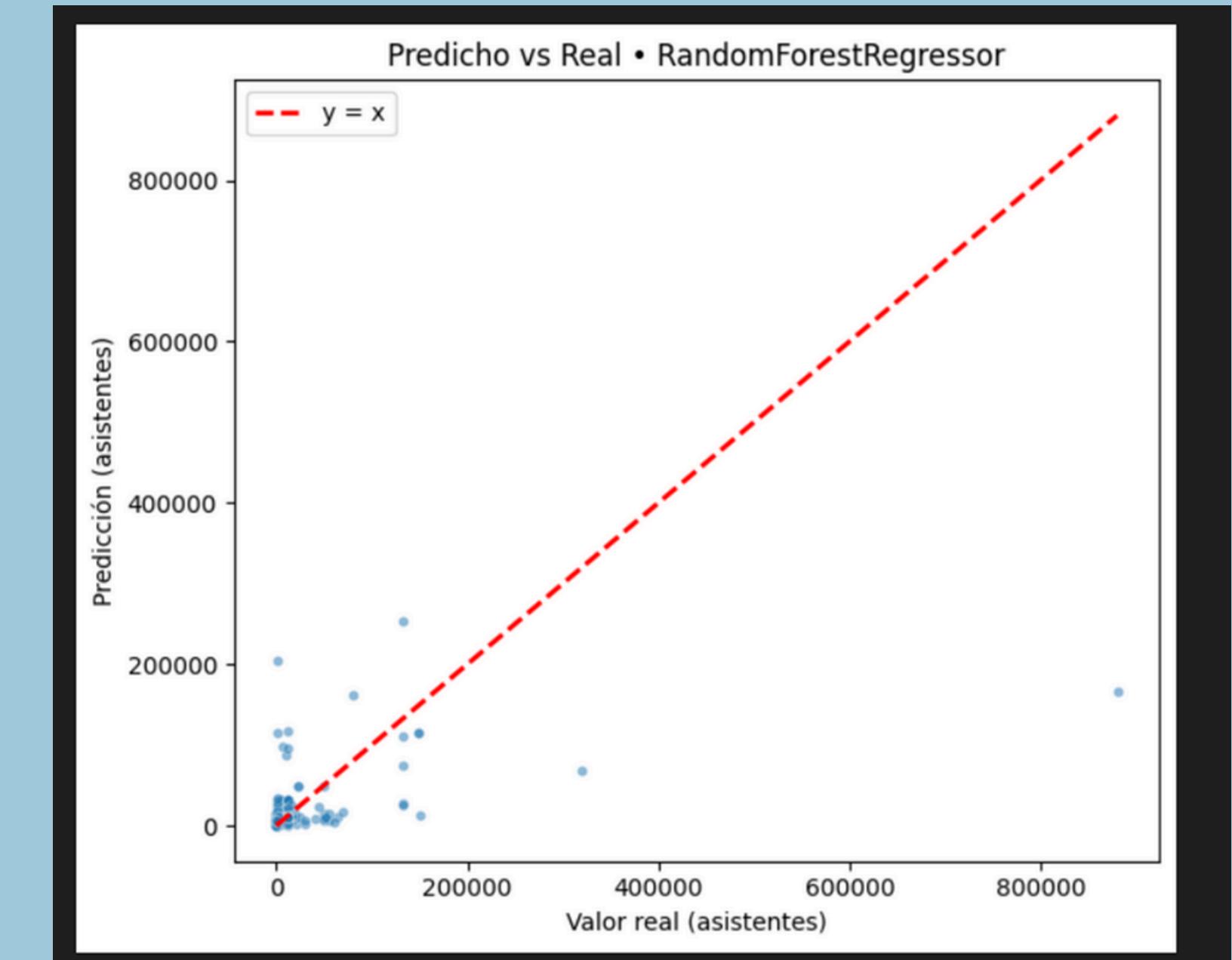
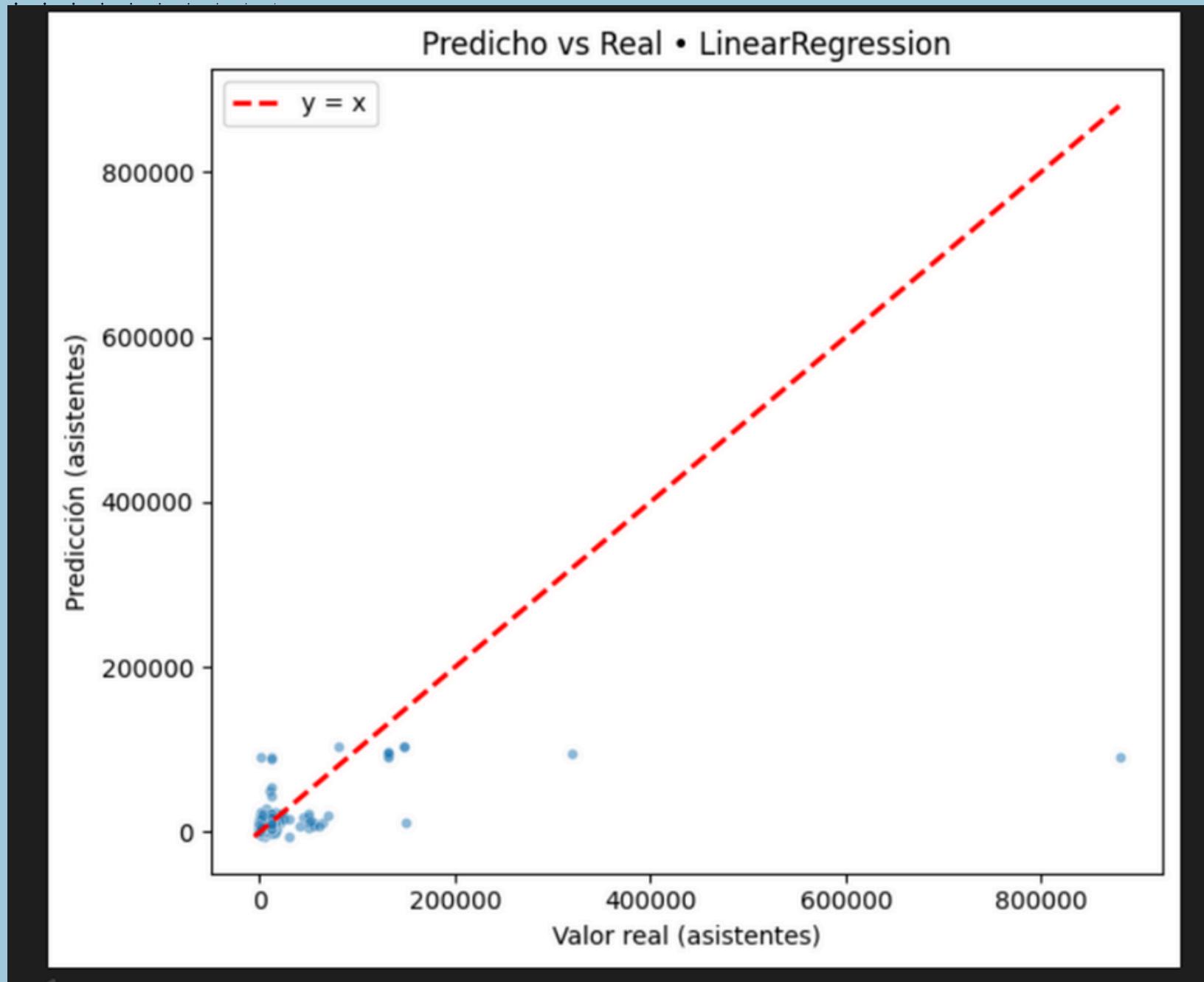
Evaluación:

- Métricas: MAE, RMSE, R².
- Validación cruzada.

	Modelo	MAE	RMSE	R ²
1	RandomForestRegressor	7195.318121	34418.889903	0.250052
0	LinearRegression	7934.725000	34470.318375	0.247809



Gráficos obtenidos

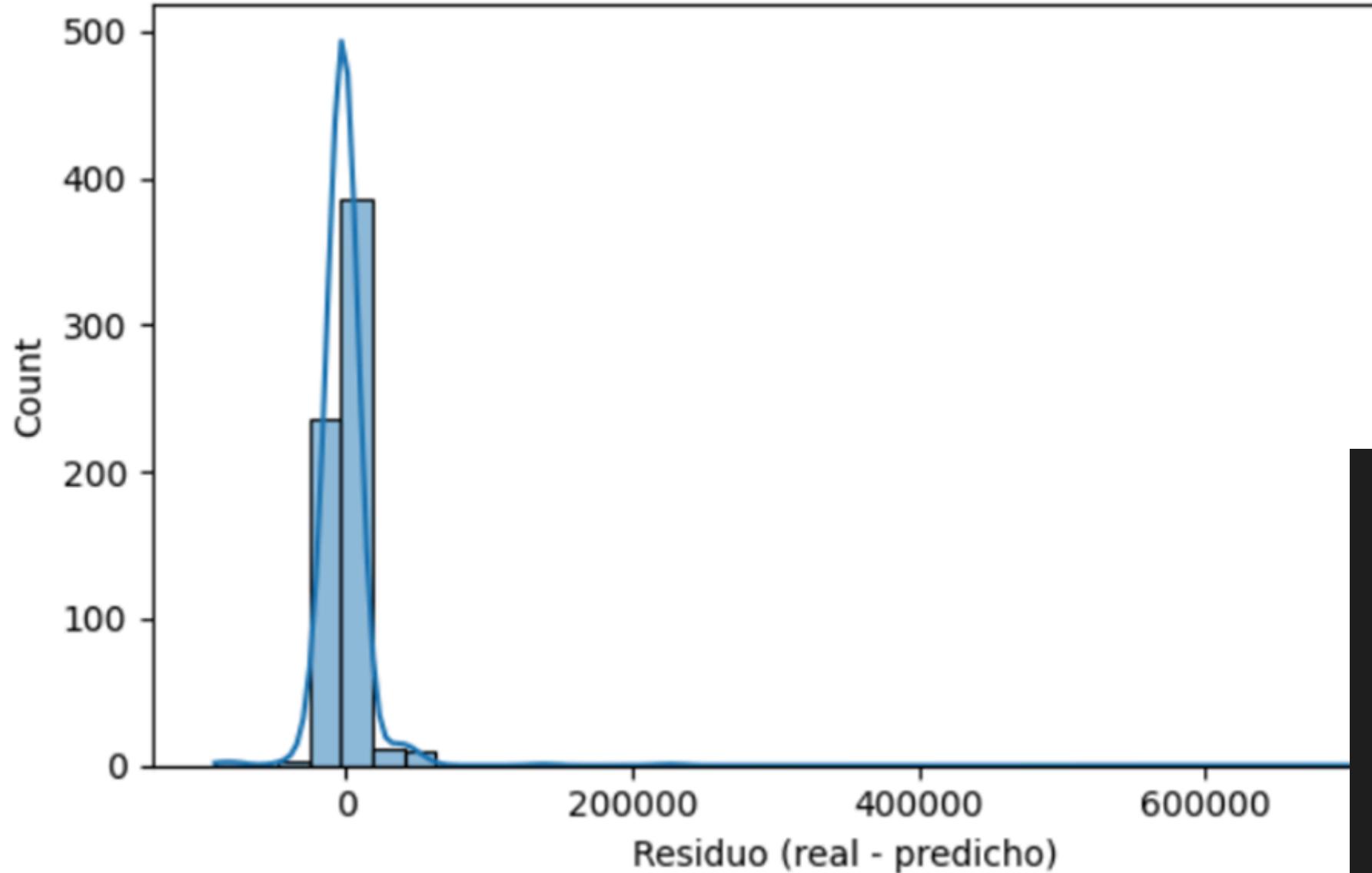


Gráficos obtenidos

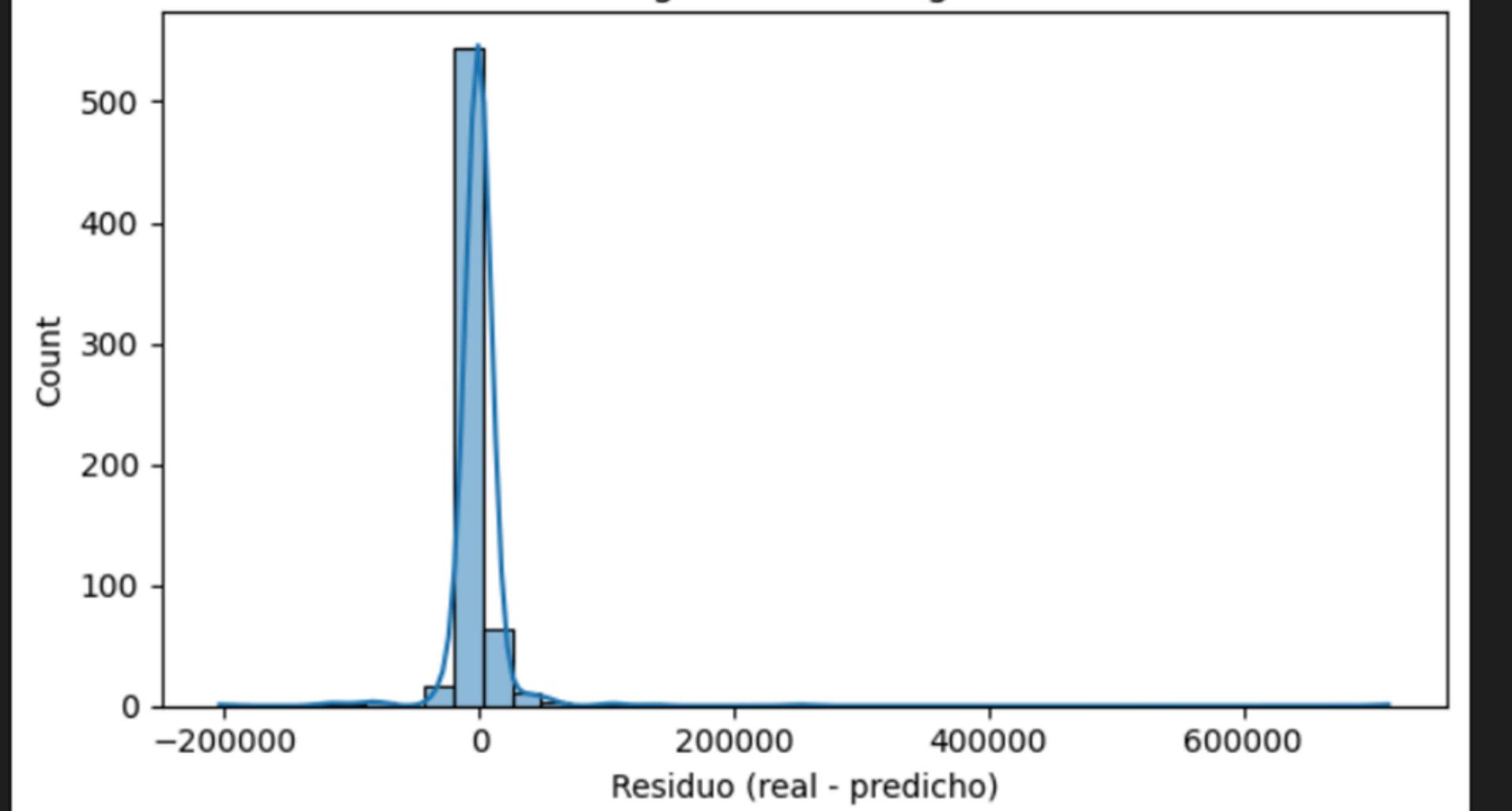


CENTRO POLÍTÉCNICO SUPERIOR
MALVINAS ARGENTINAS

LinearRegression - Histograma de residuos

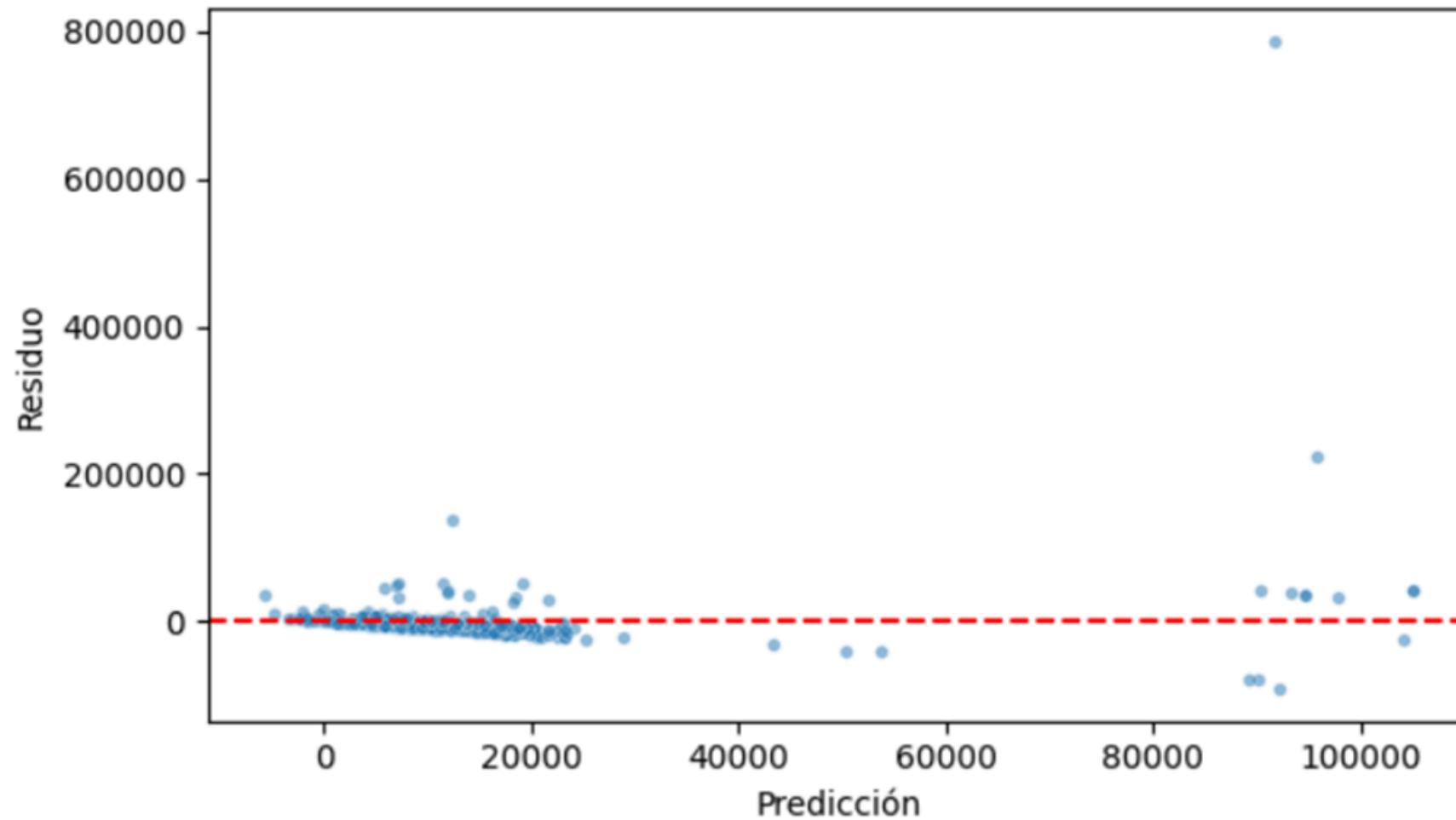


RandomForestRegressor - Histograma de residuos



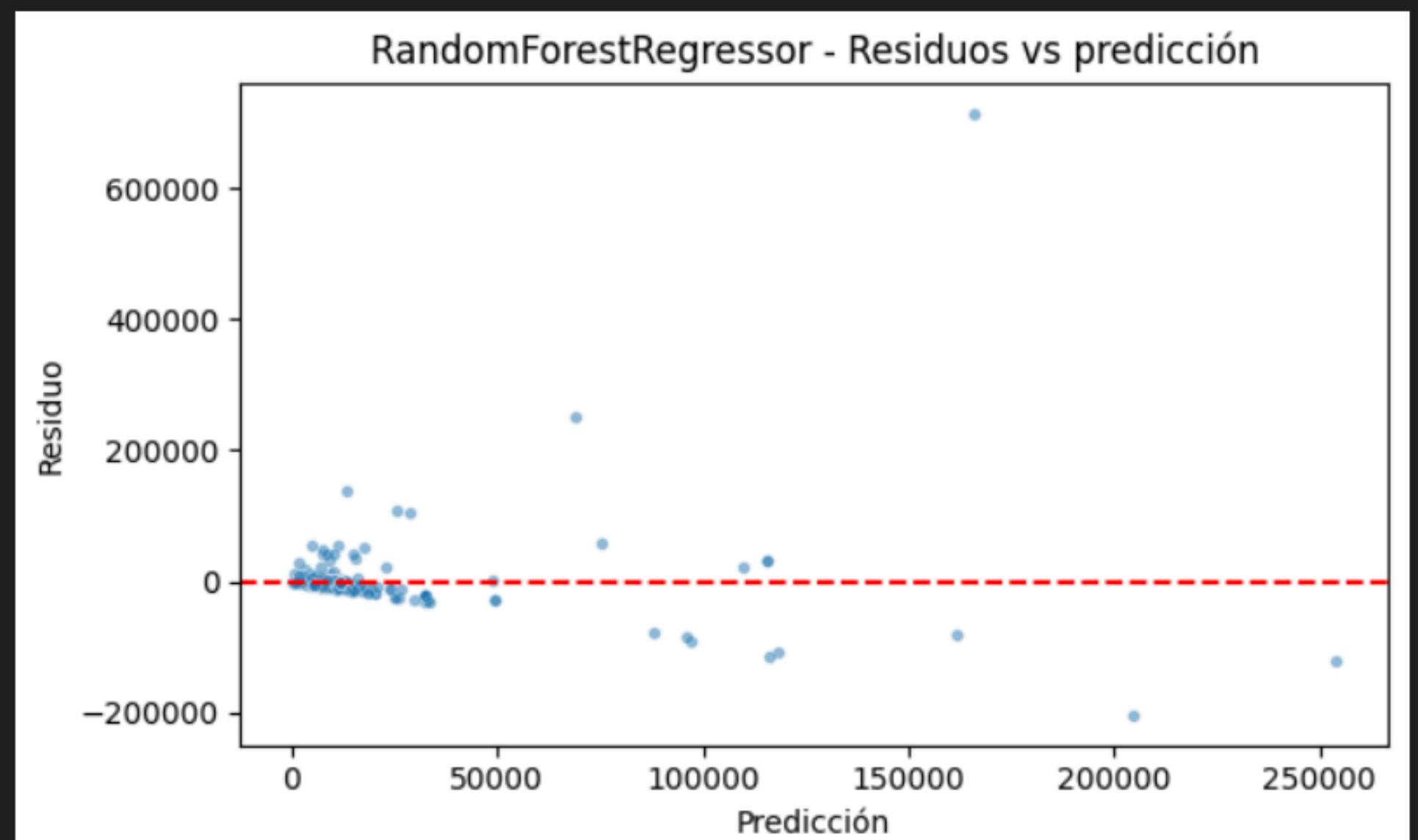


LinearRegression - Residuos vs predicción



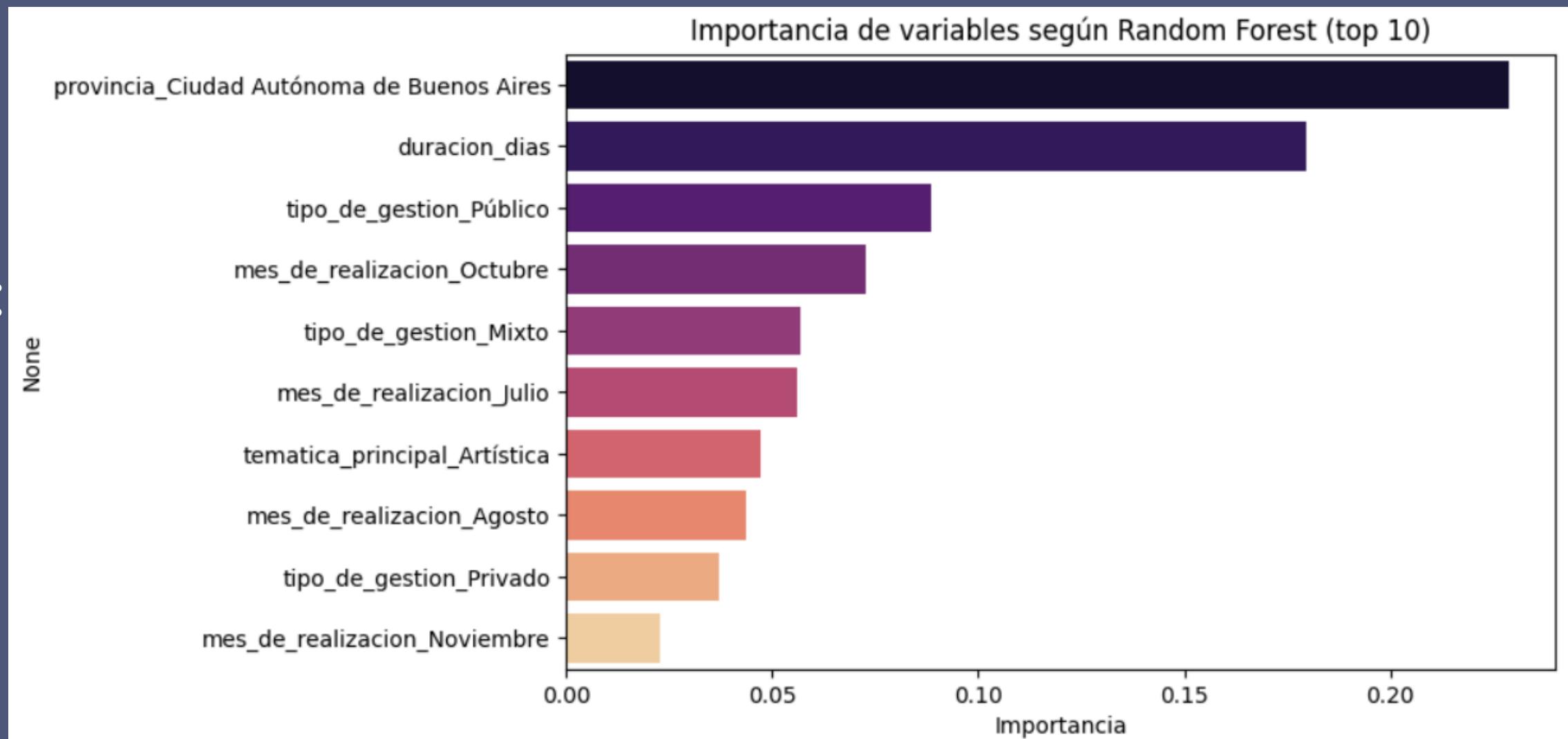
Gráficos obtenidos

RandomForestRegressor - Residuos vs predicción



Análisis y resultados principales

- Random Forest obtuvo mejor desempeño (menor error, mayor R²).
- Variables con mayor impacto:
- Duración del evento
- Tipo de gestión
- Provincia



Modelos con variable logarítmica

Para mejorar la estabilidad de los modelos, apliqué una transformación logarítmica a la cantidad aproximada de asistentes.

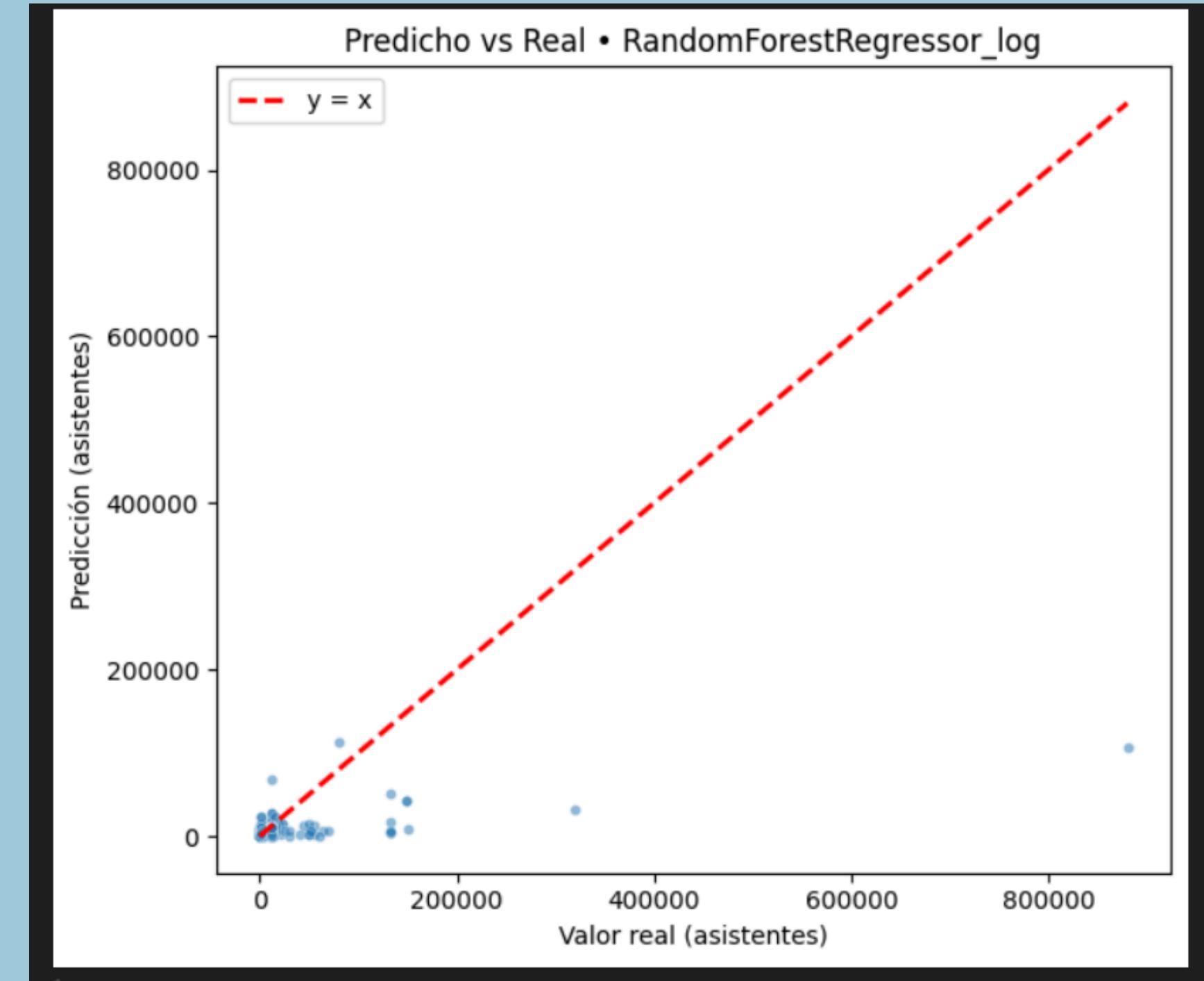
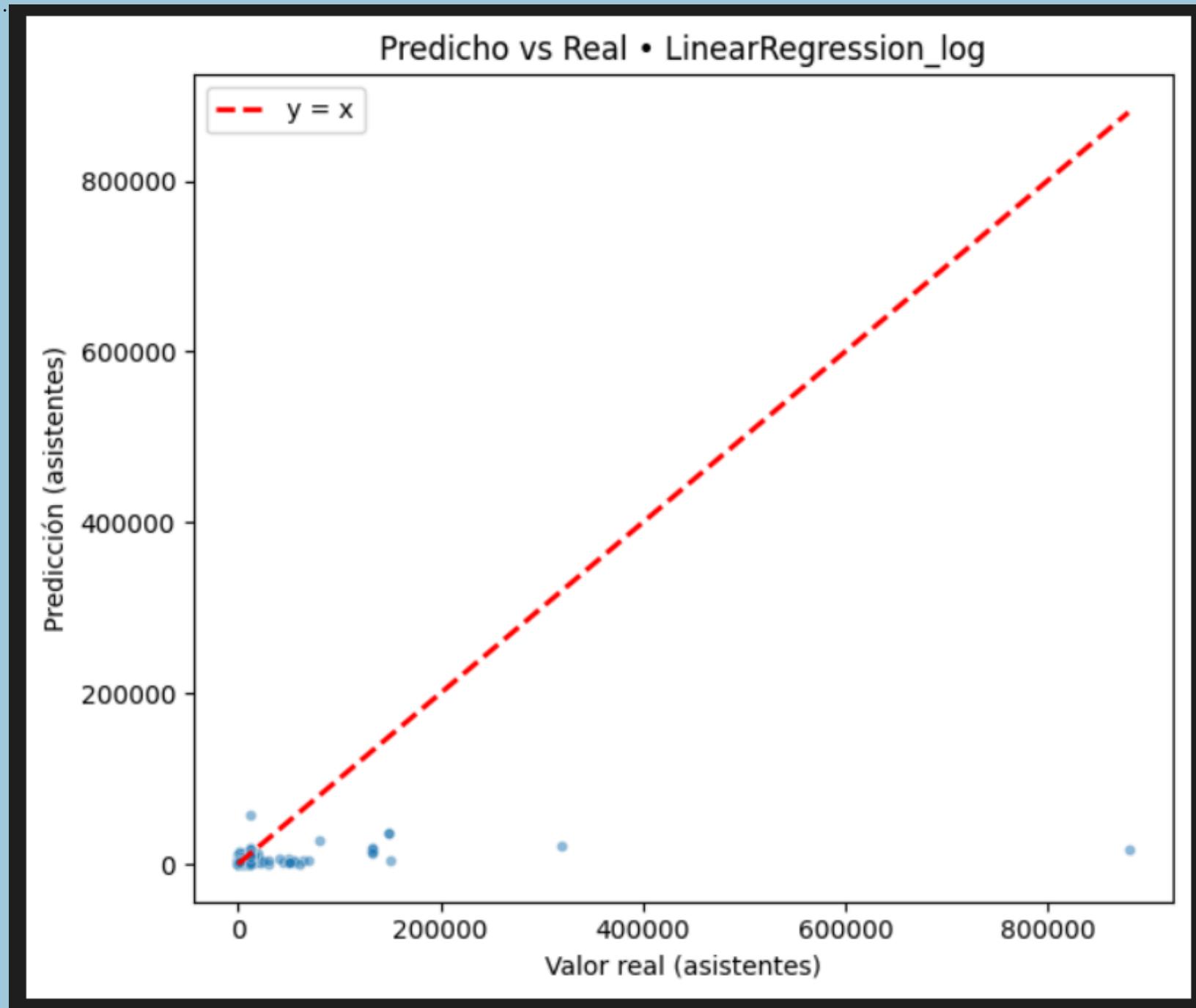
Usé `np.log1p(y)` para entrenar y luego `np.expm1()` para volver a la escala original.

Esto permitió suavizar la influencia de los festivales extremadamente grandes y equilibrar la distribución de la variable.

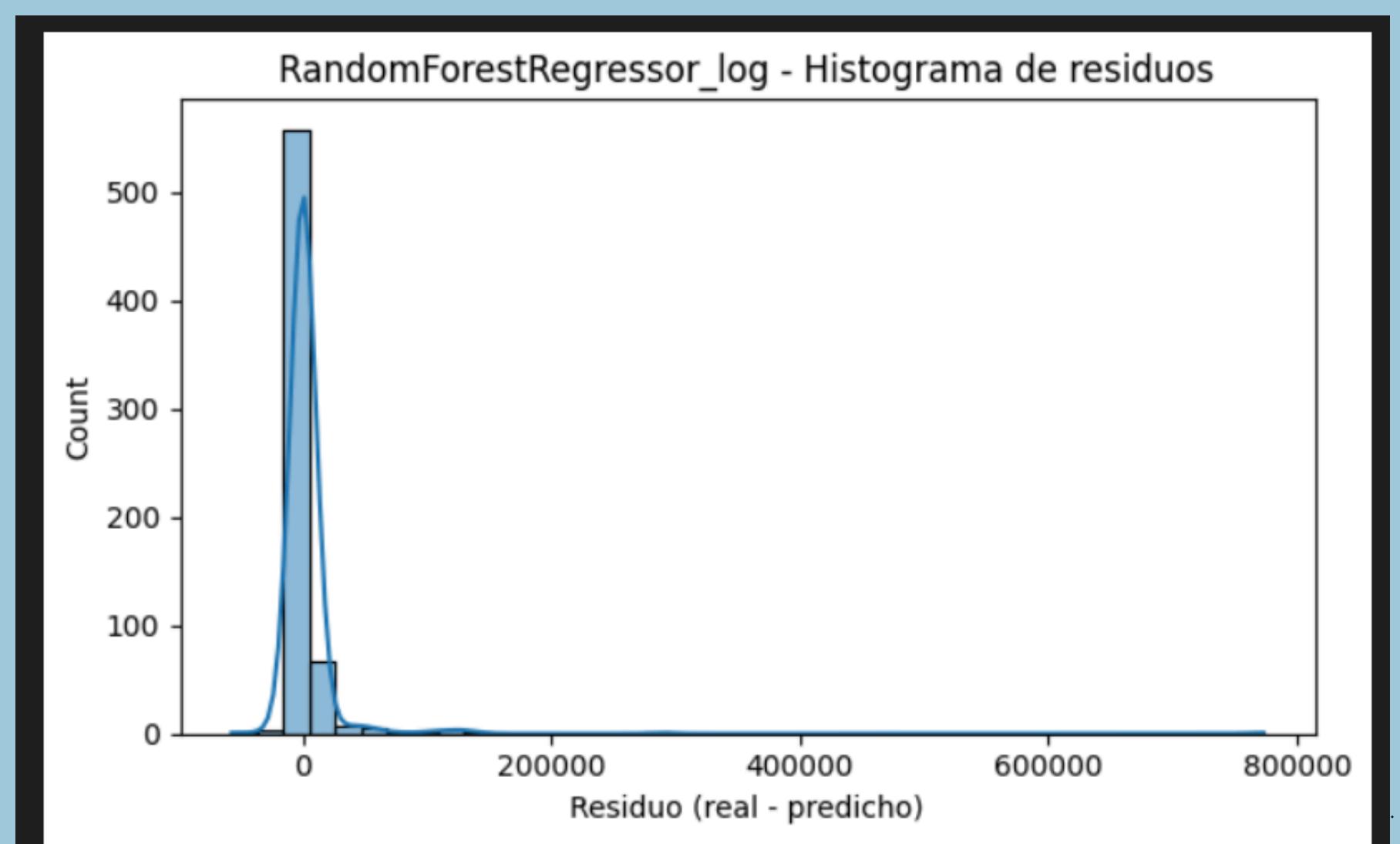
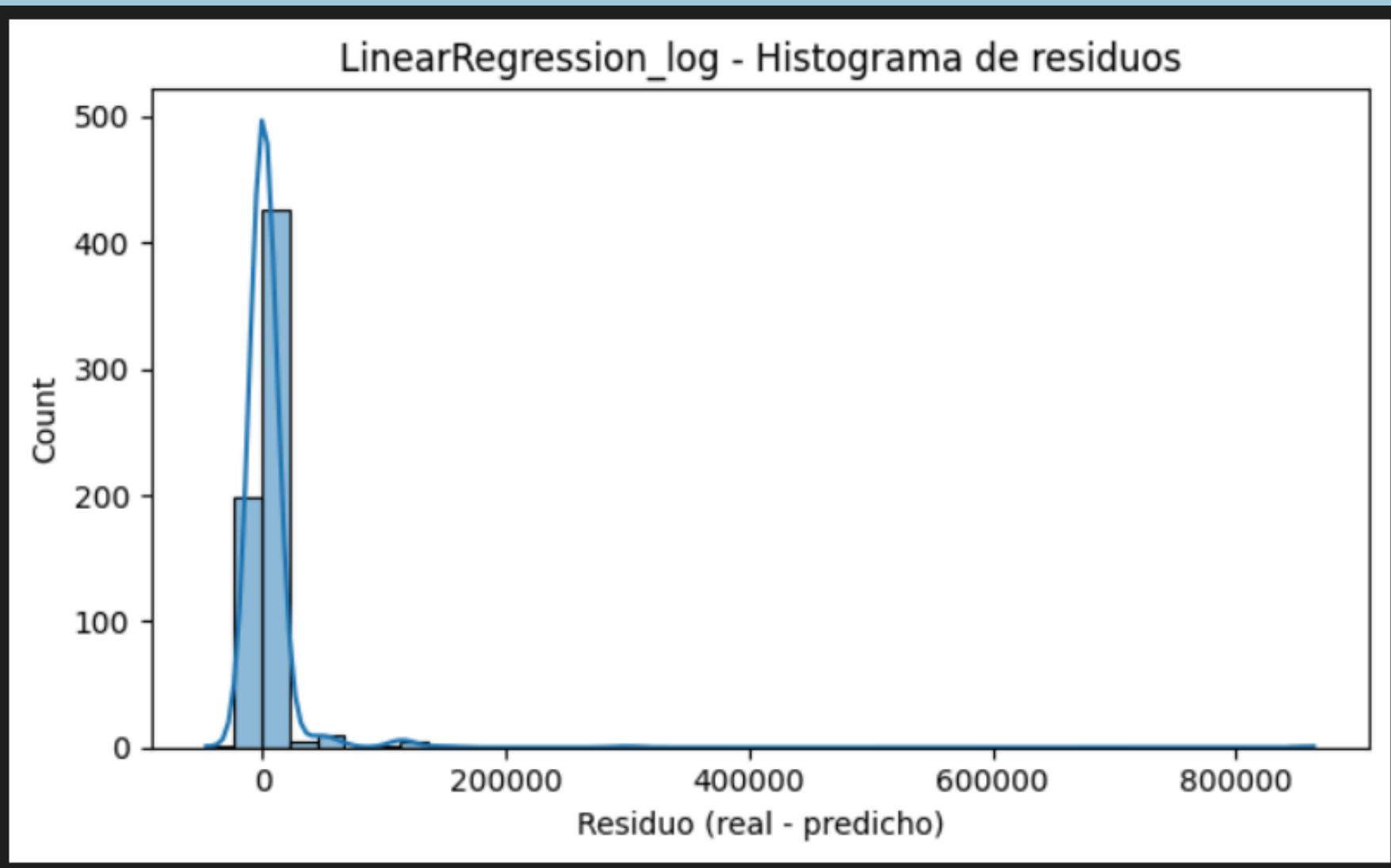
Los modelos entrenados con esta versión, llamados `LinearRegression_log` y `RandomForestRegressor_log`, mostraron una mejora notable, sobre todo en el error medio absoluto y en el RMSE.”

	Modelo	MAE	RMSE	R ²
1	RandomForestRegressor_log	6258.177432	35779.898150	0.189570
0	LinearRegression_log	7226.577095	39130.781245	0.030664

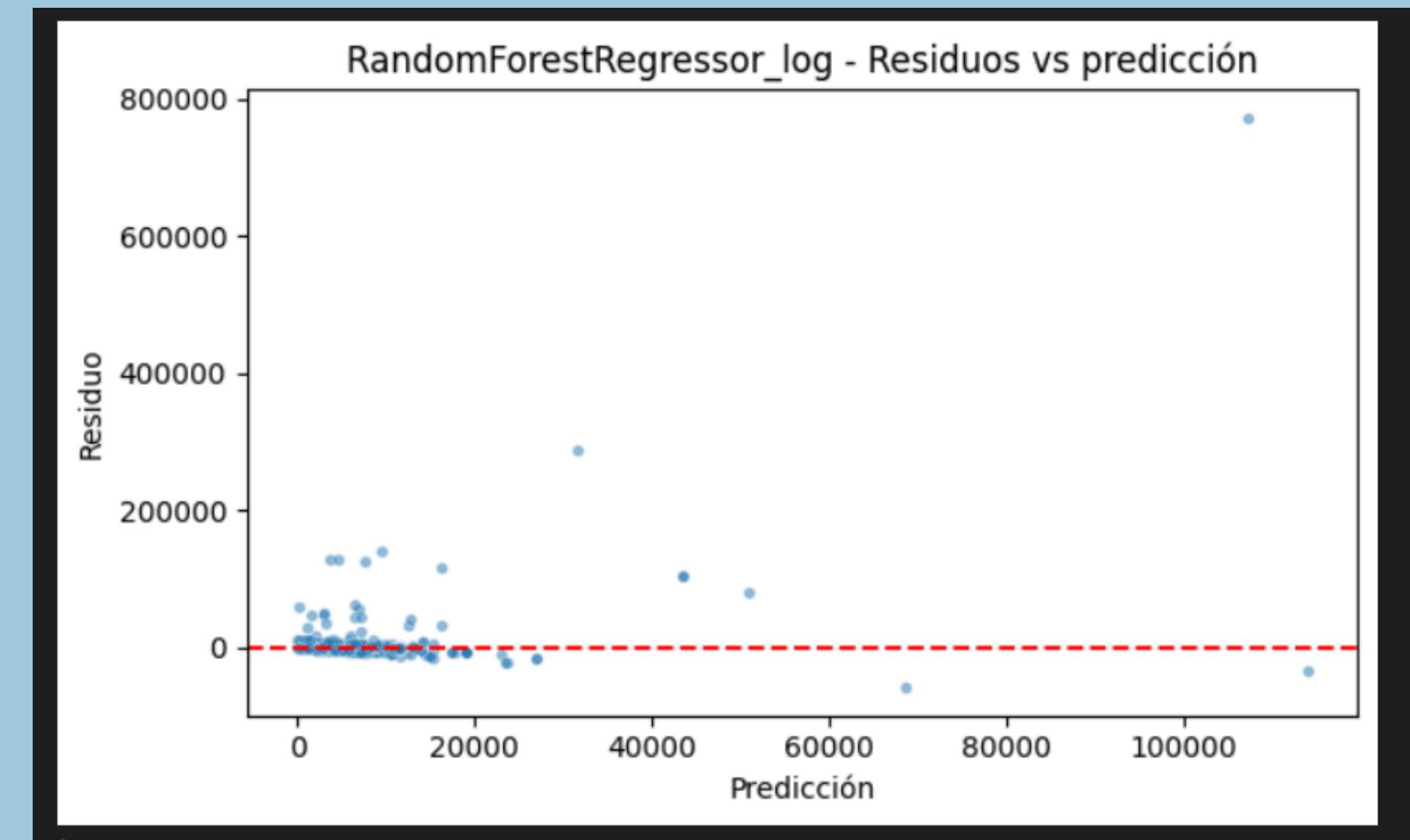
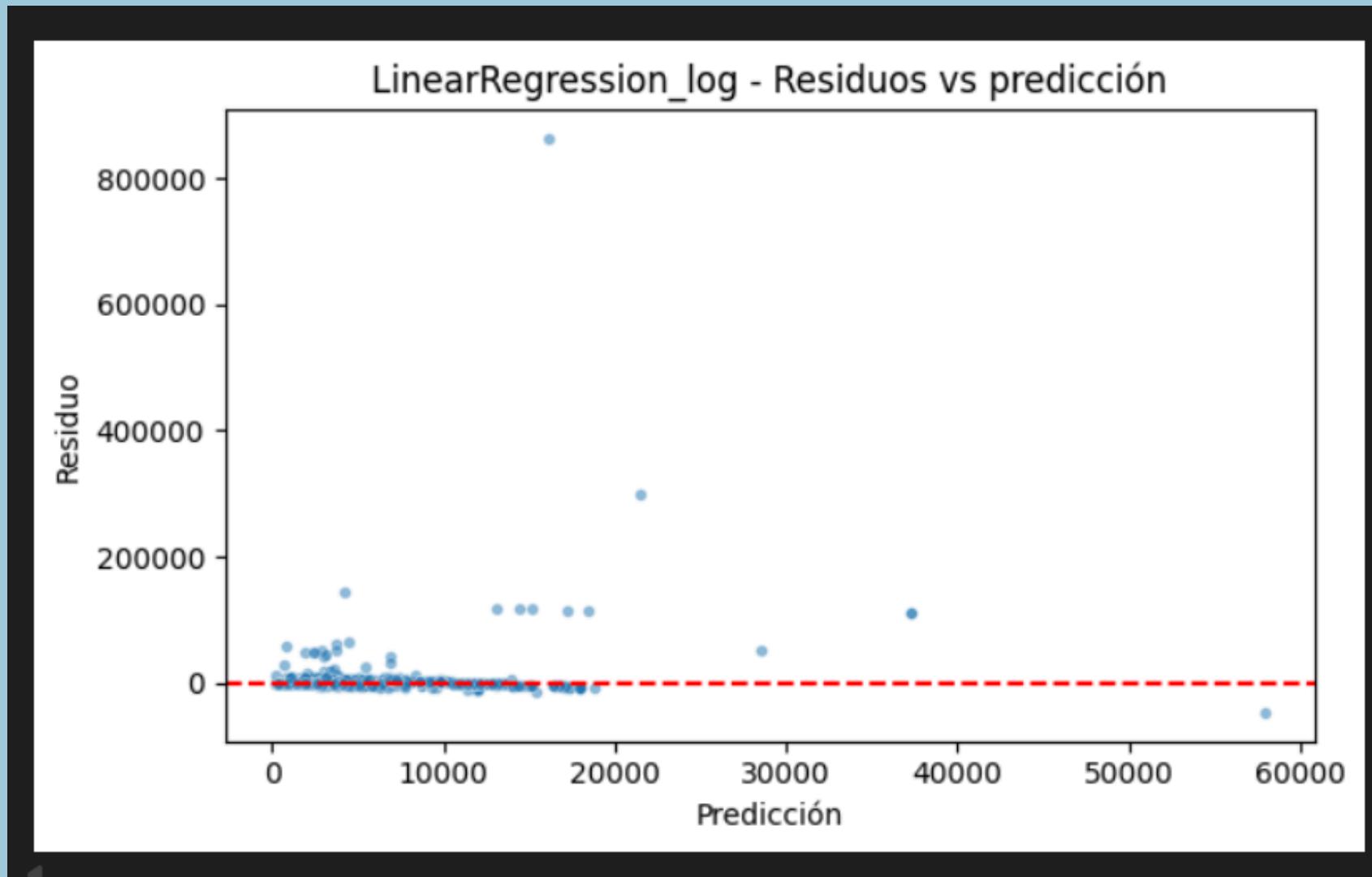
Gráficos logarítmicos obtenidos



Gráficos logarítmicos obtenidos



Gráficos logarítmicos obtenidos



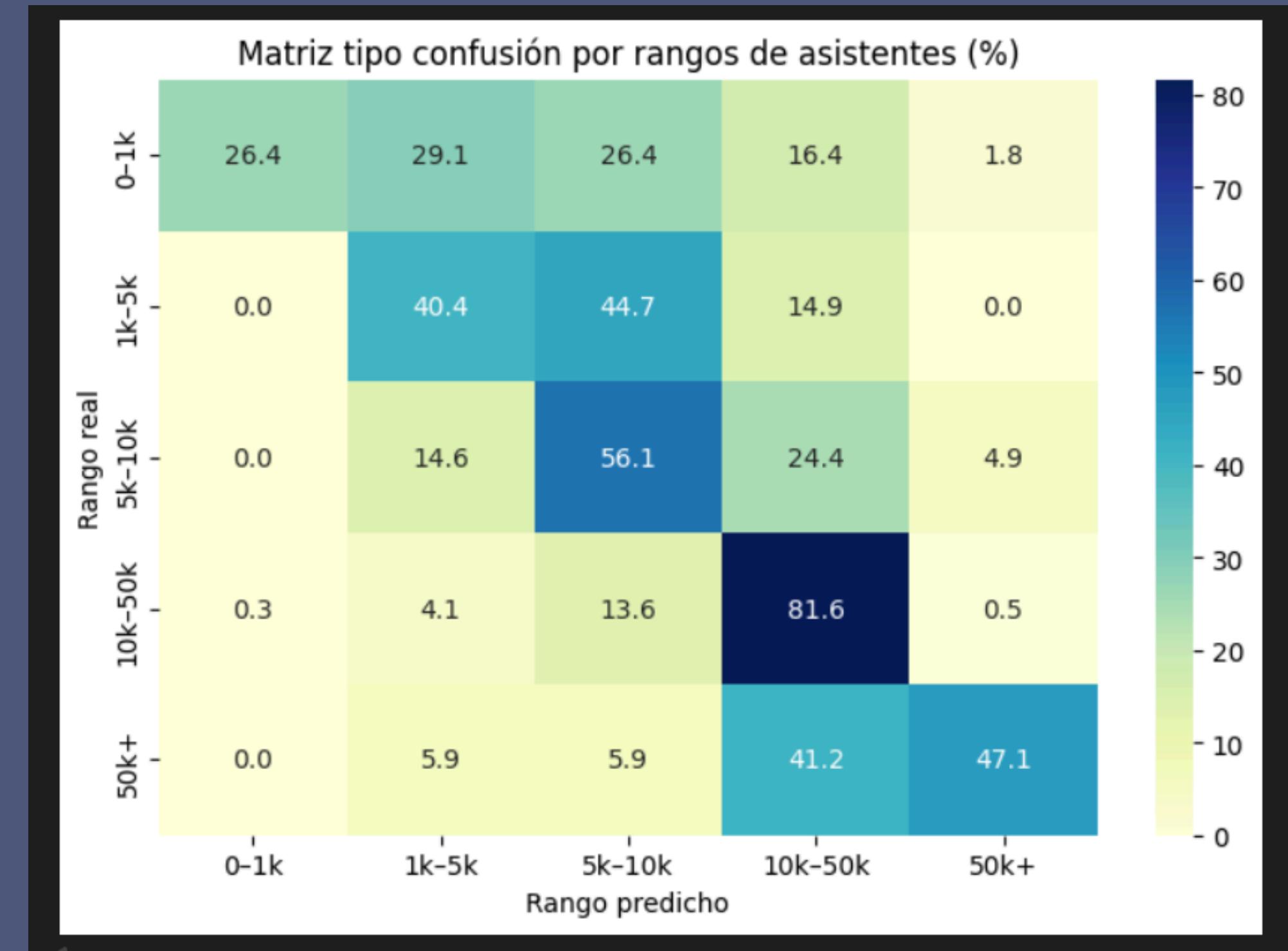


Analisis y conclusiones del modelo logaritmico

- Random Forest con variable logarítmica fue el modelo con mejor desempeño general.
- El uso del logaritmo ayudó a reducir el sesgo de los eventos masivos y a mejorar la consistencia en las predicciones.
- Las métricas mostraron una disminución del MAE y una distribución de errores más equilibrada.
- Las variables más influyentes siguieron siendo la provincia, temática principal, tipo de gestión, tipo de entrada y duración del evento

Evaluación visual del desempeño del modelo

Los resultados muestran que el modelo acierta mayormente en los rangos intermedios, con un desempeño particularmente sólido en 10k–50k, y errores que, en su mayoría, se producen hacia categorías laterales. En los extremos se observa el patrón habitual de las regresiones: moderación de máximos y mayor dispersión relativa en 0–1k.



Uso del modelo



File Edit Selection View Go Run Terminal Help

Entrega 3.ipynb • DATA S Elegí una opción (número) o Enter/'q' para cancelar: (Press 'Enter' to confirm or 'Escape' to cancel)

notebooks > Entrega 3.ipynb > M4 Asistencia promedio por provincia (Top 15) > M4 Probador interactivo de predicciones > import numpy as np

OPEN EDITORS 1 unsaved

Welcome Codificacion_dat... Entrega 3.ipynb... DATA SET INPUT... festivales_codific... metrics.csv mod... DATA SET CODIF...

Generate + Code + Markdown | Run All Restart Clear All Outputs Jupyter Variables Outline ...

.venv (3.13.6) (Python 3.13.6)

```
Modelo cargado: best_model_log.pkl
Post-proceso: expm1 (predicción en log revertida)
columnas tomadas de X.columns

Probador por consola – Predicción de asistentes
Escribí Enter o 'q' en cualquier paso para salir.

Provincia
1. Buenos Aires
2. Catamarca
3. Chaco
4. Chubut
5. Ciudad Autónoma de Buenos Aires
6. Corrientes
7. Córdoba
8. Entre Ríos
9. Formosa
10. Jujuy
11. La Pampa
12. La Rioja
13. Mendoza
14. Misiones
15. Neuquén
...
21. Santa Fe
22. Santiago del Estero
23. Tierra del Fuego
24. Tucumán
```

Output is truncated. View as a scrollable element or open in a text editor. Adjust cell output settings...

OUTLINE TIMELINE

Ln 1, Col 1 (1311 selected) Spaces: 4 Spaces: 4 LF () Finish Setup Cell 49 of 49

Conclusiones finales

- El modelo logra una buena capacidad predictiva en los rangos intermedios (entre 10 k y 50 k asistentes), donde se concentra la mayoría de los eventos.
- Se observan mayores desvíos en los extremos –festivales muy pequeños o masivos–, lo cual es un comportamiento esperable en modelos de regresión con datos heterogéneos.
- Las variables más influyentes resultaron ser la duración del evento, la provincia y el tipo de gestión.
- En conjunto, el modelo permite estimar la asistencia con una precisión aceptable y demuestra cómo la ciencia de datos puede contribuir a la planificación cultural y turística basada en evidencia.



**Muchas gracias
por la atención**



**CENTRO POLÍTÉCNICO SUPERIOR
MALVINAS ARGENTINAS**