



# INFORME

*de Negocios*



# PARTE I: CASO DE NEGOCIO A RESOLVER



El banco estima que **retener a un cliente** tiene un costo promedio de **40 USD** por campaña de retención, mientras que **adquirir uno nuevo** cuesta en promedio **200 USD**.

Además, un **cliente activo** con tarjeta genera en promedio **120 USD** de margen anual.

## Datos:

Clientes iniciales: 20,000

Cada año se adquiere +5% (200 USD / nuevo)

Cada año se van -2% (churn)

## Objetivo del proyecto a 10 años:

Diseñar un sistema de predicción de fuga por cliente con rendimiento objetivo de precisión  $\approx 70\%$ .

## 2) Preguntas clave que un modelo de predicción debería responder

### PROBABILIDAD DE FUGA ANUALIZADA

¿Cuál es la probabilidad de abandono en cada año 1-10?

### EVOLUCIÓN DE VARIABLES PREDICTORAS

¿Cómo cambian demográficas, transaccionales y de comportamiento a lo largo del tiempo y su impacto en la fuga?

### IMPACTO DE INTERVENCIÓN AÑO A AÑO

¿Cuánto reduce la intervención de retención la probabilidad de fuga en cada año, considerando el costo de \$40 y el beneficio esperado?

### PUNTAJE DE RIESGO

¿Qué clientes presentan alto riesgo acumulado de fuga a lo largo de la década?

### VENTANAS DE OBSERVACIÓN COMPARABLES

¿Qué rendimiento se obtiene al usar historiales de 3, 6 y 12 meses para actualizar predicciones anuales?

## **COSTOS Y BENEFICIOS A LARGO PLAZO**

- ROI anual y acumulado de campañas de retención.
- VPN/LTV descontado al año 0, con y sin intervención, usando tasas de descuento del 3% al 10%.

## **INCORPORACIÓN DE NUEVOS PRODUCTOS**

¿Cómo afecta la adición de productos al CLV y a la probabilidad de fuga?

## **SEGURIDAD Y SESGOS A LARGO PLAZO**

¿Mantiene equidad entre segmentos (género, edad, región) durante 10 años?  
¿Cómo mitigar sesgos de forma continua?

## **ROBUSTEZ ANTE CAMBIOS DE NEGOCIO**

¿Cómo se adapta el modelo ante cambios regulatorios, tasas de interés o comisiones?

### 3) Impacto esperado de un sistema de predicción de fuga en los indicadores del banco

Como hoy:

- Ingreso bruto/año: 2.4M
- Costo por adquisiciones (20% de 20k = 4,000 clientes × 200): USD 800k/año
- Neto/año: USD 1.6M → usas 33.3% del ingreso bruto solo en CAC
- Neto acumulado 10y (sin descuento): USD 16.0M

Implementación:

- Ingreso bruto/año: 2.4M (mantienes tamaño estable)
- Costo de retención (contactas ≈ 4,000/año): USD 160k/año
- CAC residual (repones 1,200/año): USD 240k/año
- Costos totales/año: USD 400k → 16.7% del ingreso bruto
- Neto/año: USD 2.0M (+25% vs Escenario A)
- Neto acumulado 10y (sin descuento): USD 20.0M

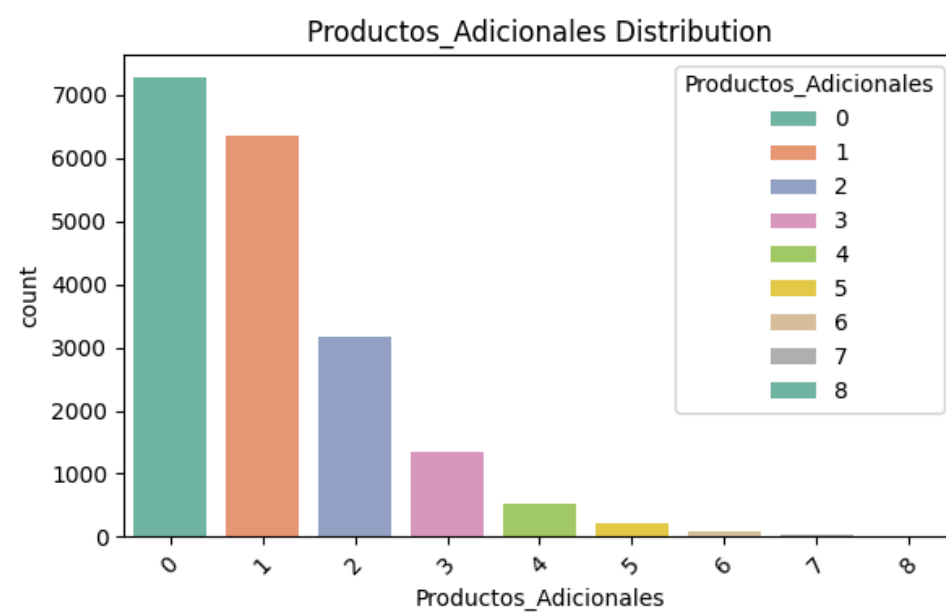
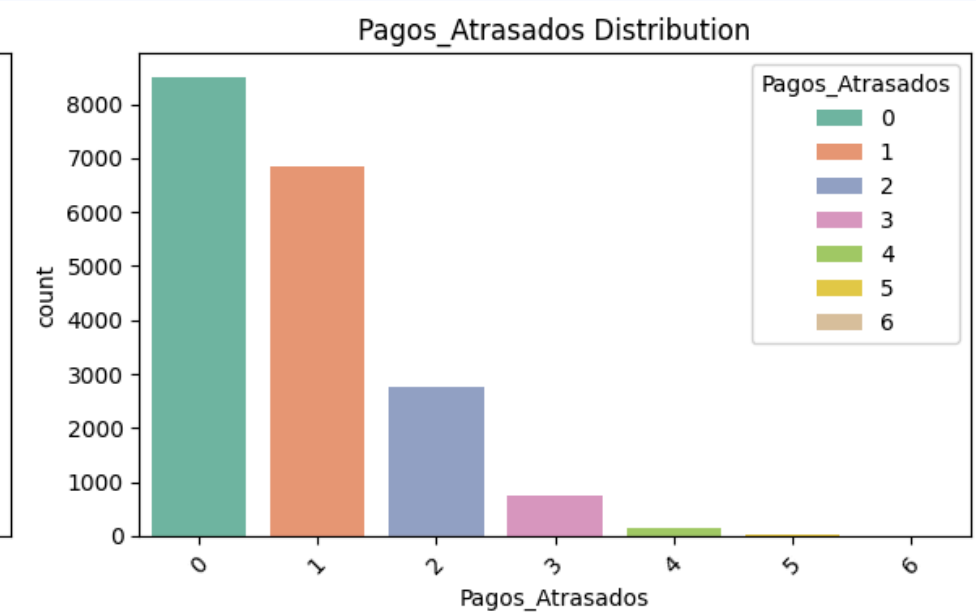
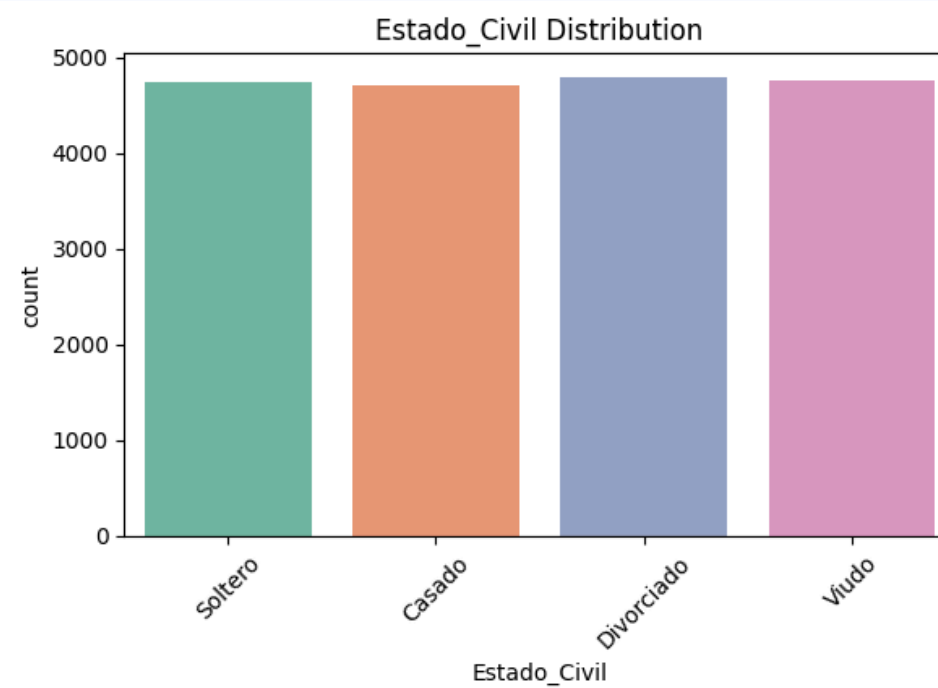
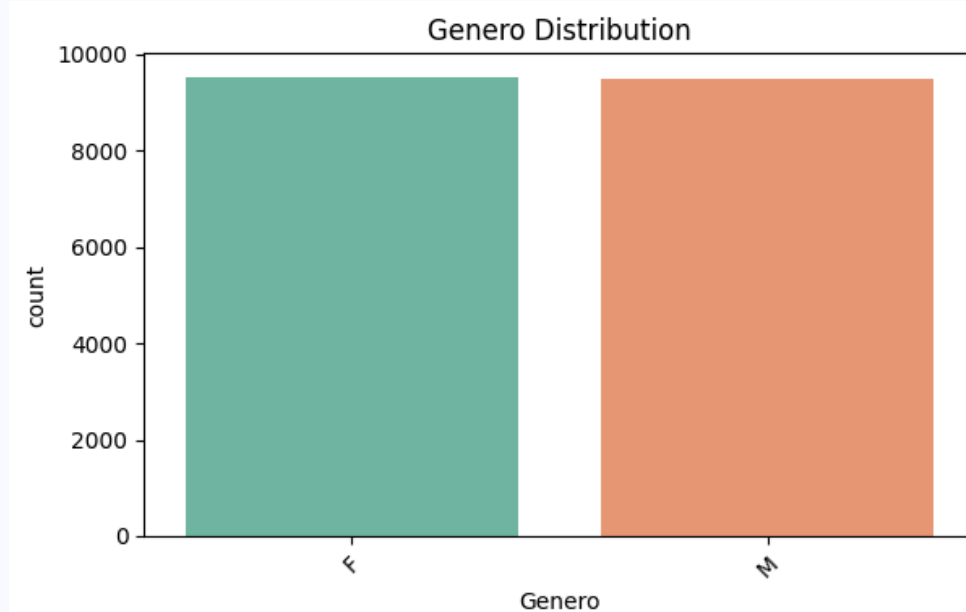
No haces nada:

- Ingreso 10y acumulado (por decaimiento 20%): USD 10.71M
- Neto/año: decreciente; sin costos (ni CAC ni retención)
- % vs "ideal sin fuga" (24.0M): 44.6% (pierdes 55.4%)

Métricas de implementación

- Retención: 70%
- Gasto de equipo x 10 años: 200k
- Ingreso esperado 10 años: 422k
- Margen 10 años: 222k
- ROI: 110%

# PARTE 2 : ANÁLISIS DE LOS DATOS PARA POSIBLE EXPLOTACIÓN



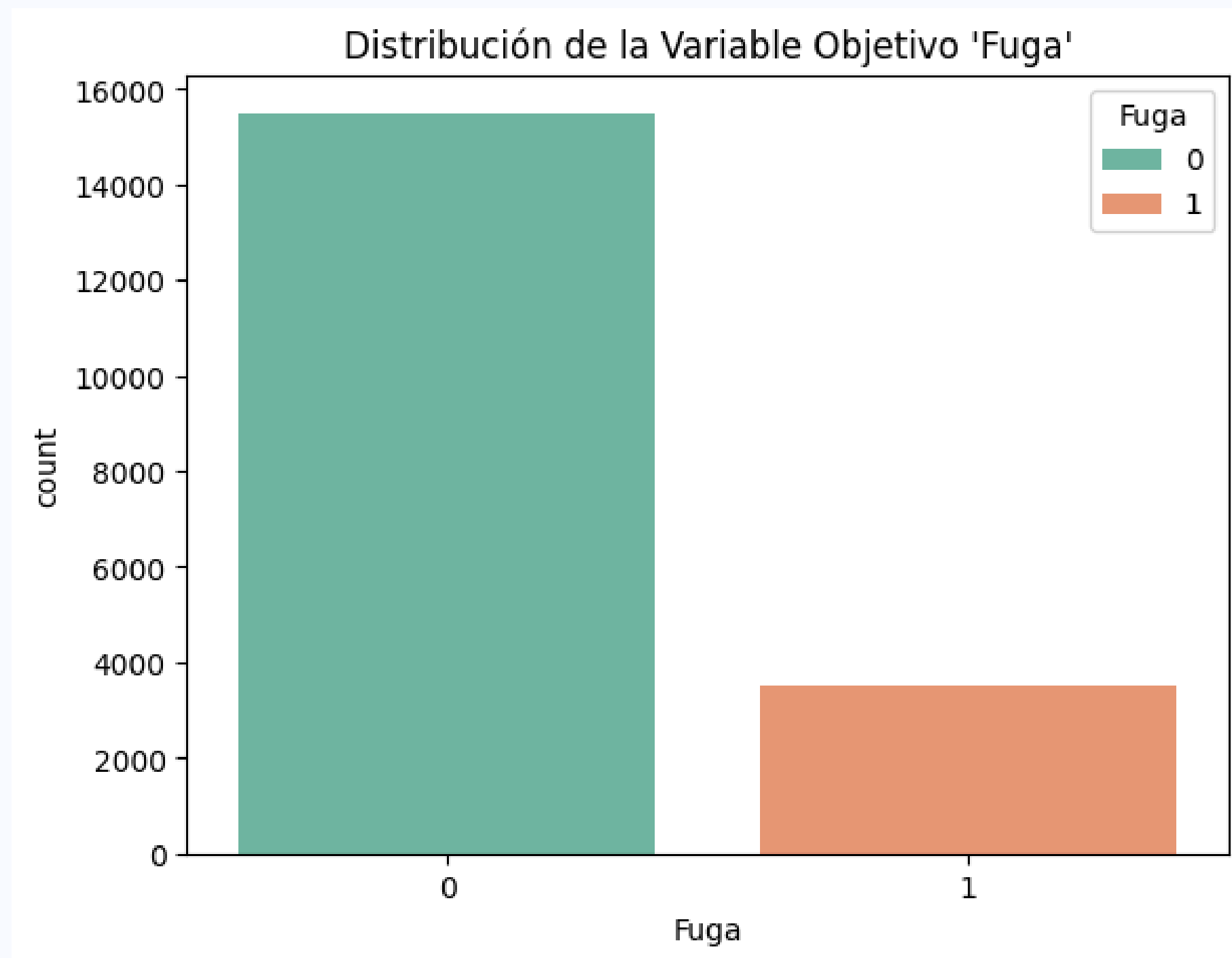
**Género:** Muestra balanceada

**Estado\_Civil:** Muestra balanceada

**Pagos\_Atrasados:** La mayoría mantiene un buen historial

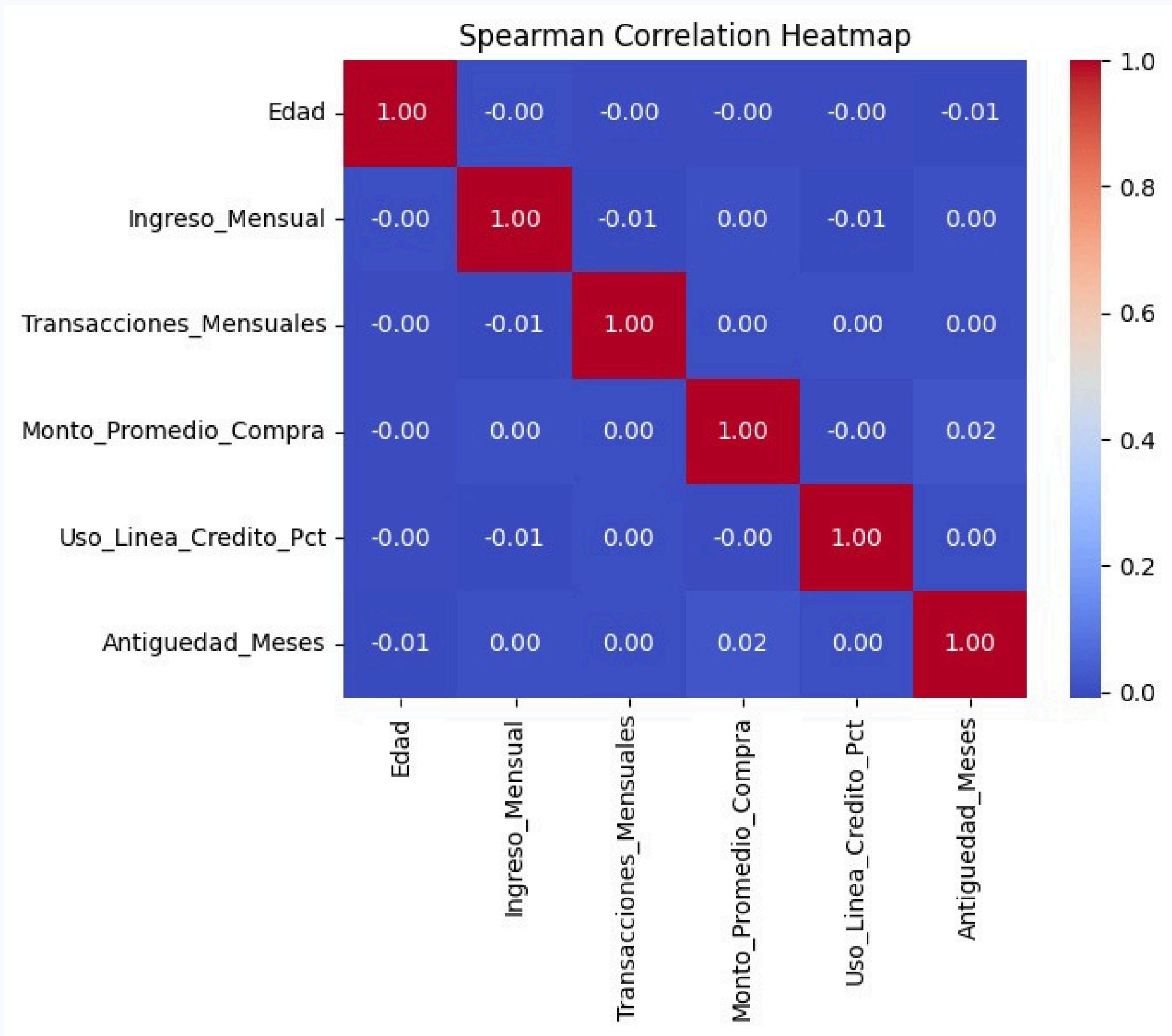
**Productos\_Adicionales:** Poca vinculación con productos adicionales

# DISTRIBUCIÓN FUGA



La variable objetivo está desbalanceada, lo que puede afectar el rendimiento de los modelos de clasificación.

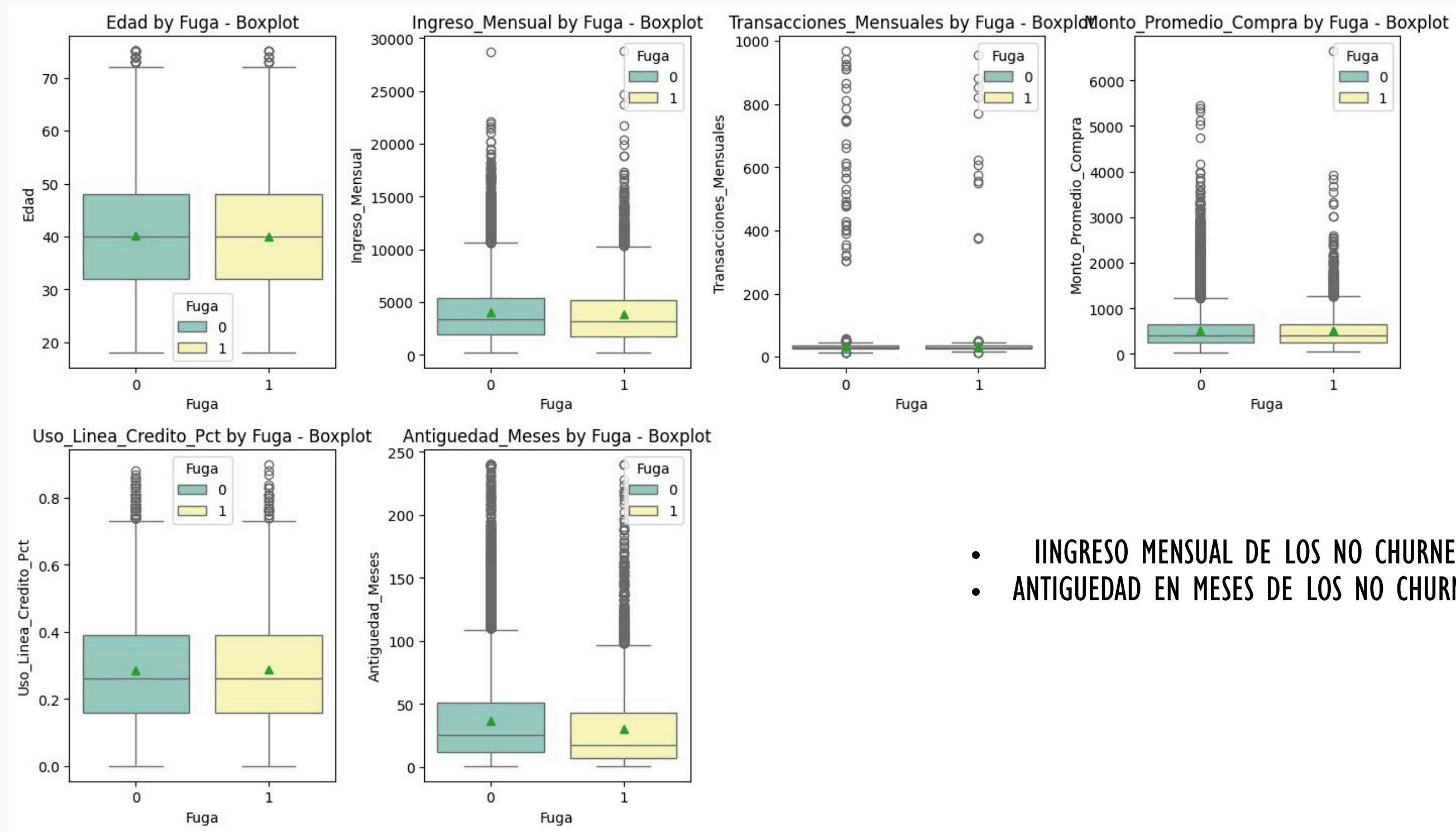
# MULTIVARIADO CUANTITATIVAS VS CUANTITATIVAS



No existen correlaciones significativas entre las variables numéricas; todos los coeficientes están muy cercanos a 0.



# MULTIVARIADO CUANTITATIVAS VS CUALITATIVAS



- INGRESO MENSUAL DE LOS NO CHURNERS TIENDE A VALORES MÁS ALTOS
- ANTIGÜEDAD EN MESES DE LOS NO CHURNERS TIENDE A VALORES MÁS ALTOS

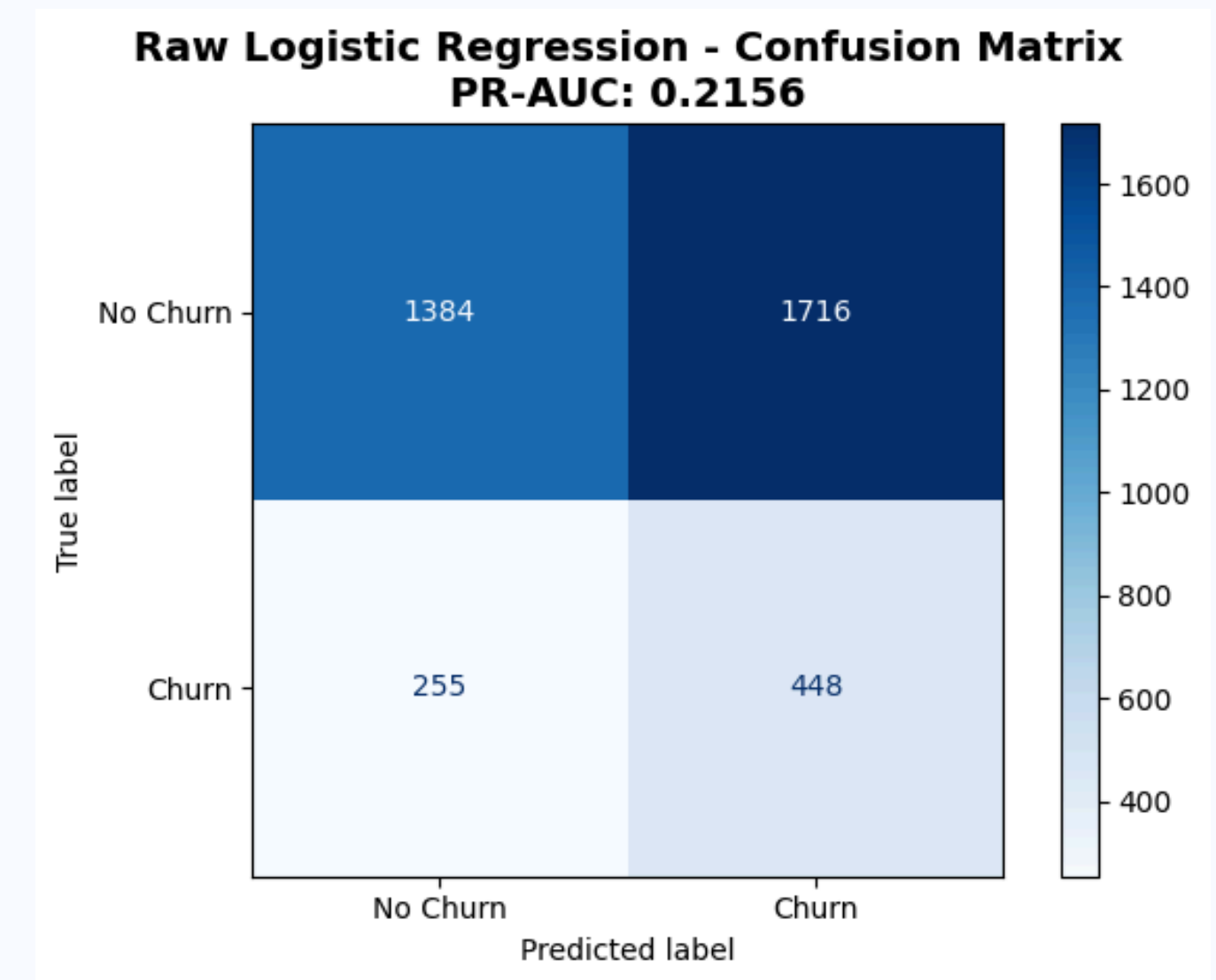
# MULTIVARIADO CUALITATIVA VS CUALITATIVA

Fuga	0	1
Pagos_Atrasados		
0	81.3	18.7
1	82.3	17.7
2	82.3	17.7
3	78.0	22.0
4	63.1	36.9
5	82.6	17.4
6	50.0	50.0

Fuga	0	1
Productos_Adicionales		
0	79.6	20.4
1	81.9	18.1
2	83.4	16.6
3	83.1	16.9
4	84.5	15.5
5	86.2	13.8
6	86.7	13.3
7	73.1	26.9
8	84.6	15.4

# MÉTRICAS BASE LOGIT DATA CRUDA, WEIGHT='BALANCED', LBGFS

accuracy: 0.6172155555571				
	precision	recall	f1-score	support
0	0.84	0.45	0.58	3100
1	0.21	0.64	0.31	703
accuracy			0.48	3803
macro avg	0.53	0.54	0.45	3803
weighted avg	0.73	0.48	0.53	3803



# FEATURE ENGINEERING

gasto\_mensual  
spend\_income  
ingreso\_residual  
tenure\_bin  
ing\_bin  
util\_band

Flags respecto a atrasos

# ESCALADO Y TRANSFORMACIONES

transformaciones logarítmicas

Robust Scaler

Clipping

# ESTRATEGIAS DE REGULARIZACIÓN LOGIT

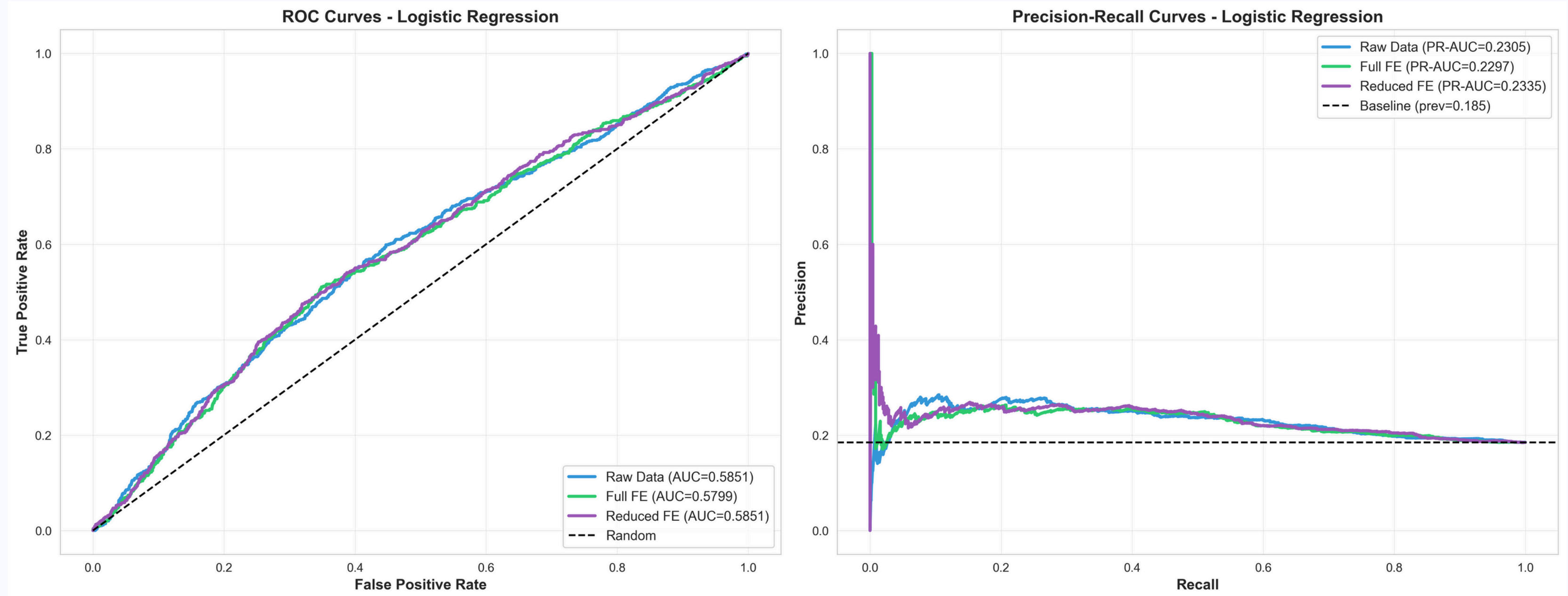
=====						
Strategy	Precision	Recall	F1	Avg Precision	Non-zero Coefs	
Current (L2, C=1.0)	0.245602	0.516358	0.332875	0.229720	60	
Stronger L2 (C=0.1)	0.245065	0.512091	0.331492	0.230178	60	
Stronger L2 (C=0.01)	0.250534	0.500711	0.333966	0.233236	60	
L1 Lasso (C=1.0)	0.245935	0.516358	0.333180	0.229899	52	
L1 Lasso (C=0.1)	0.249473	0.504979	0.333960	0.231037	39	
ElasticNet (C=0.1)	0.249307	0.512091	0.335352	0.231129	43	
Best by Average Precision:						
Strategy	Stronger L2 (C=0.01)					
Precision	0.250534					
Recall	0.500711					
F1	0.333966					
Avg Precision	0.233236					
Non-zero Coefs	60					

# FEATURE SELECTION

- Manual (VIF, CORR)
- Lasso
- L2 fuerte

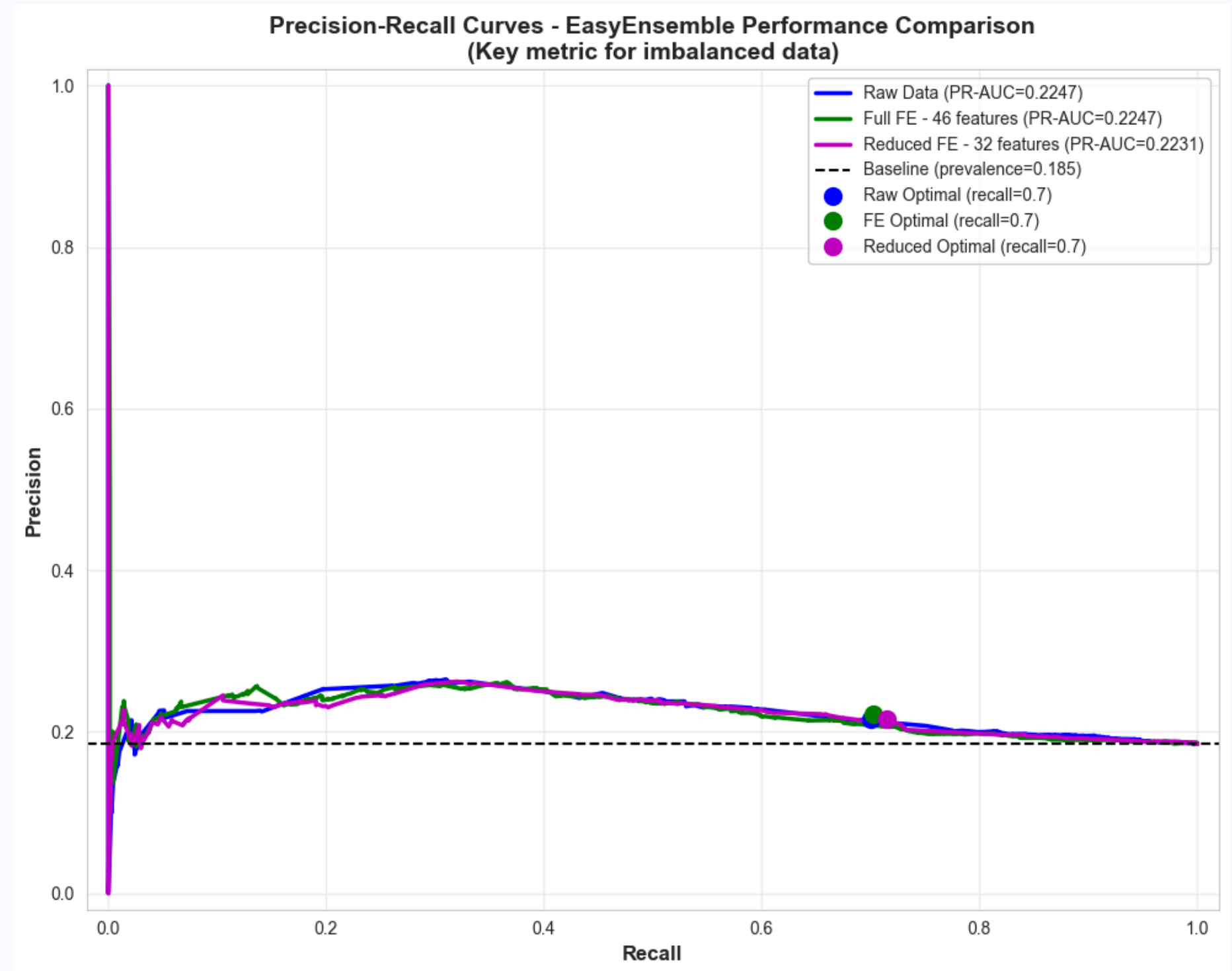
Escogimos Manual

# ROC PR CURVES LOGISTIC

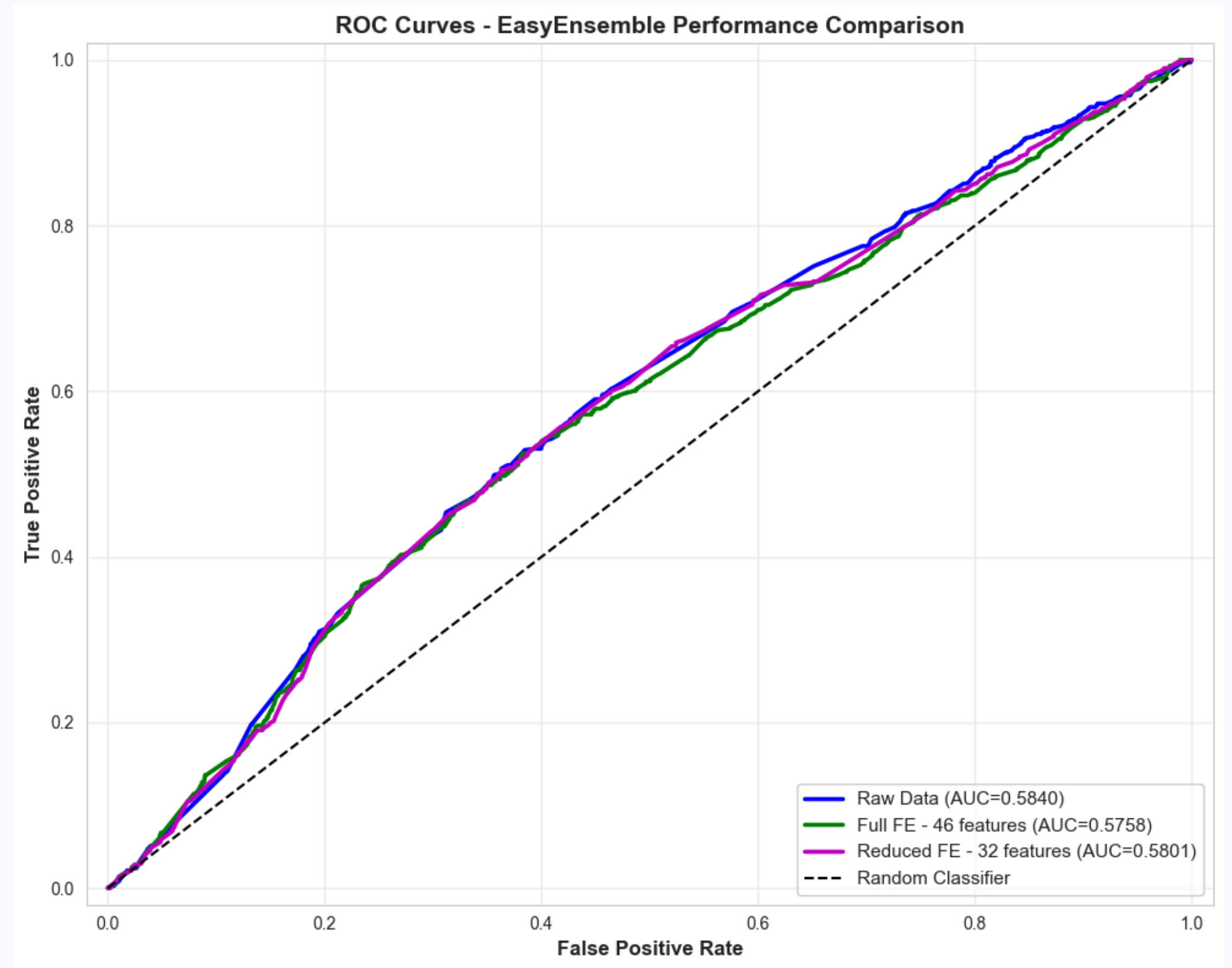




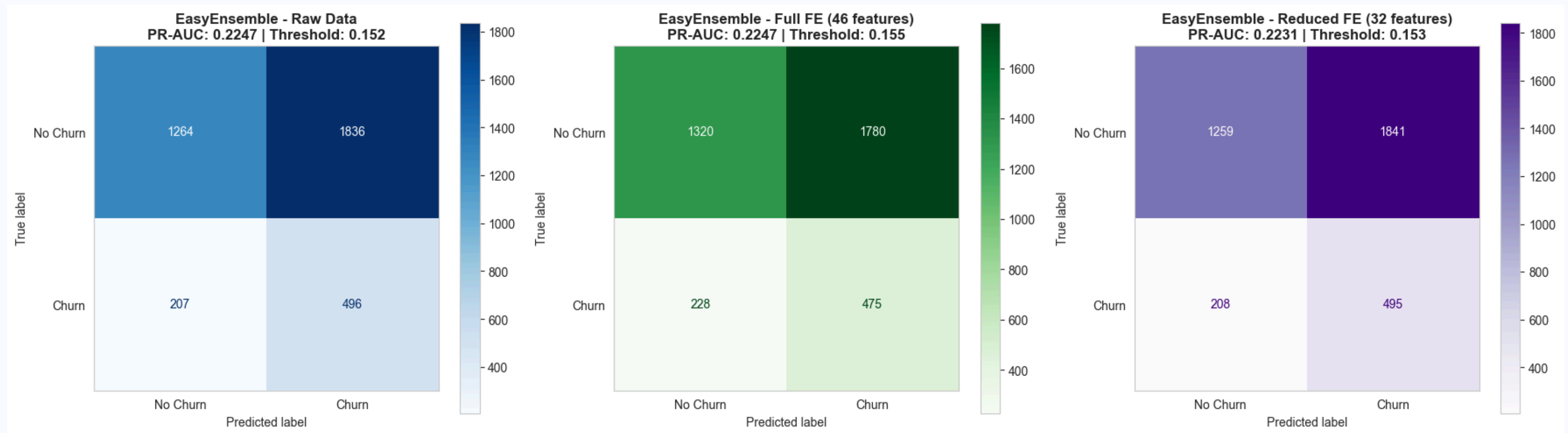
# CURVES EASYENSEMBLE



# ROC CURVES EASYENSEMBLE



# CONFUSION MATRICES EASYENSEMBLE

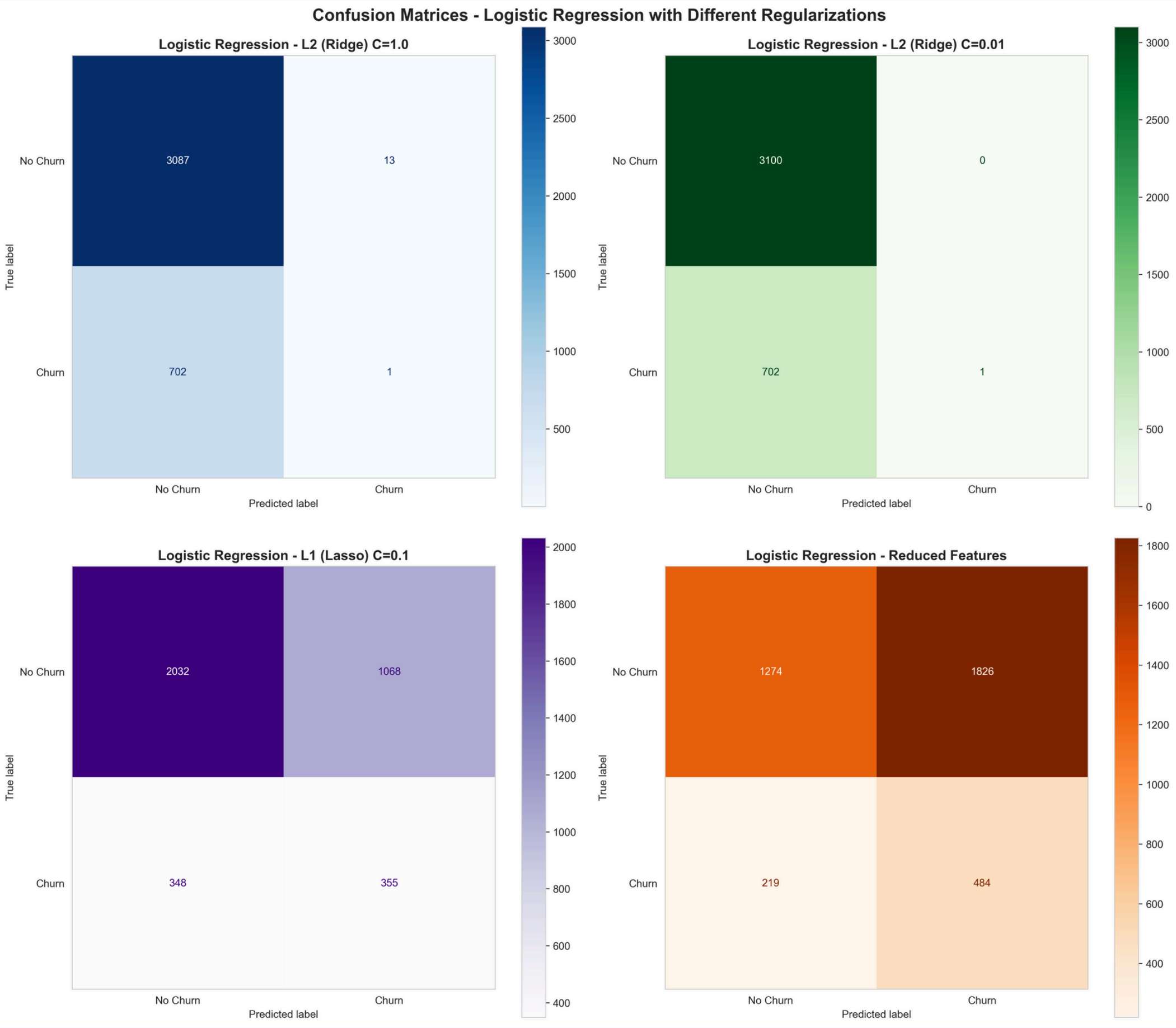


El modelo detecta una parte razonable de los clientes que sí se van (496), pero genera **muchos falsos positivos** (1836), es decir, marca a muchos clientes fieles como posibles fugas.

Modelo más conservador, que reduce falsas alarmas a costa de no detectar algunas fugas reales.

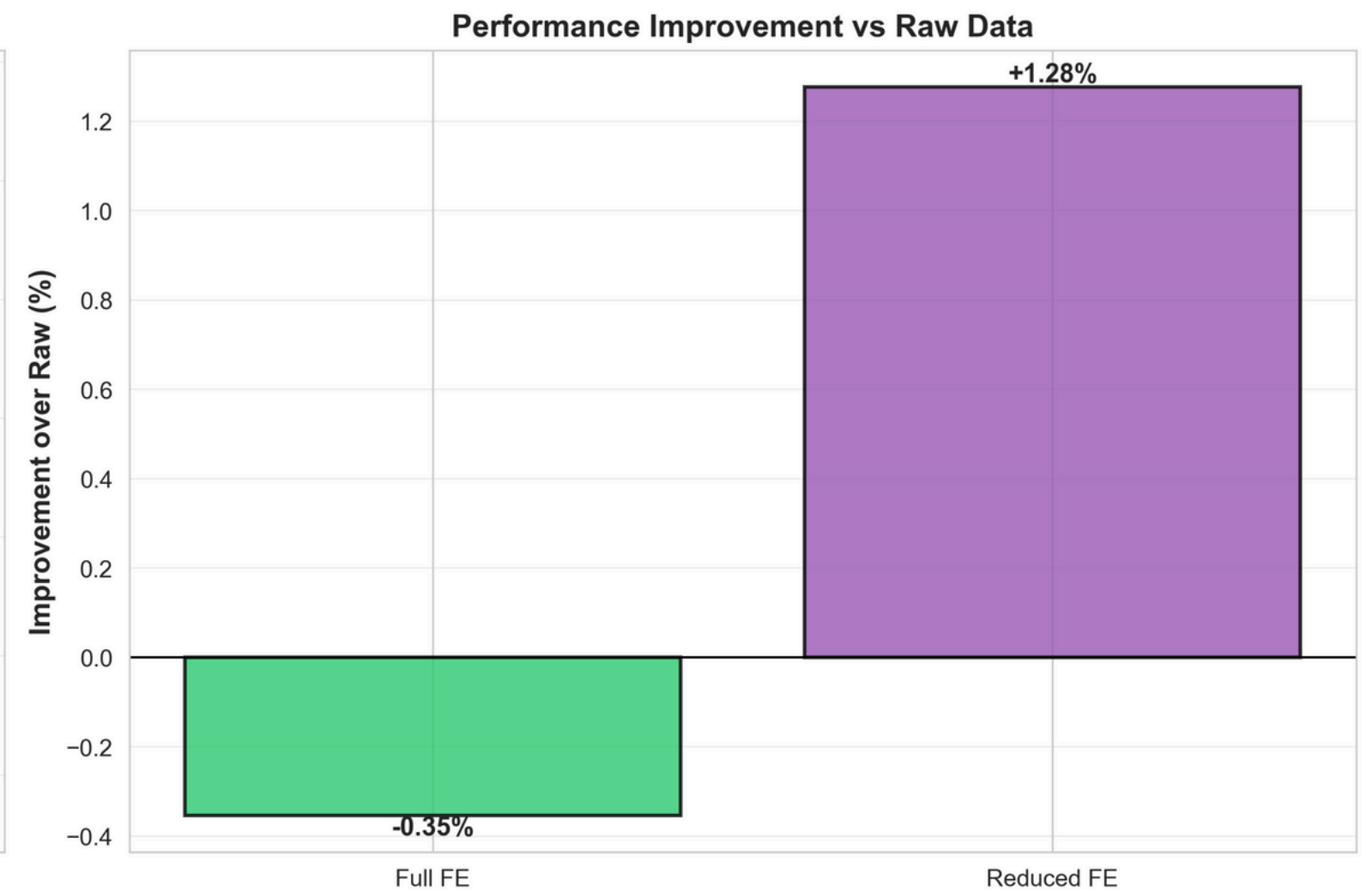
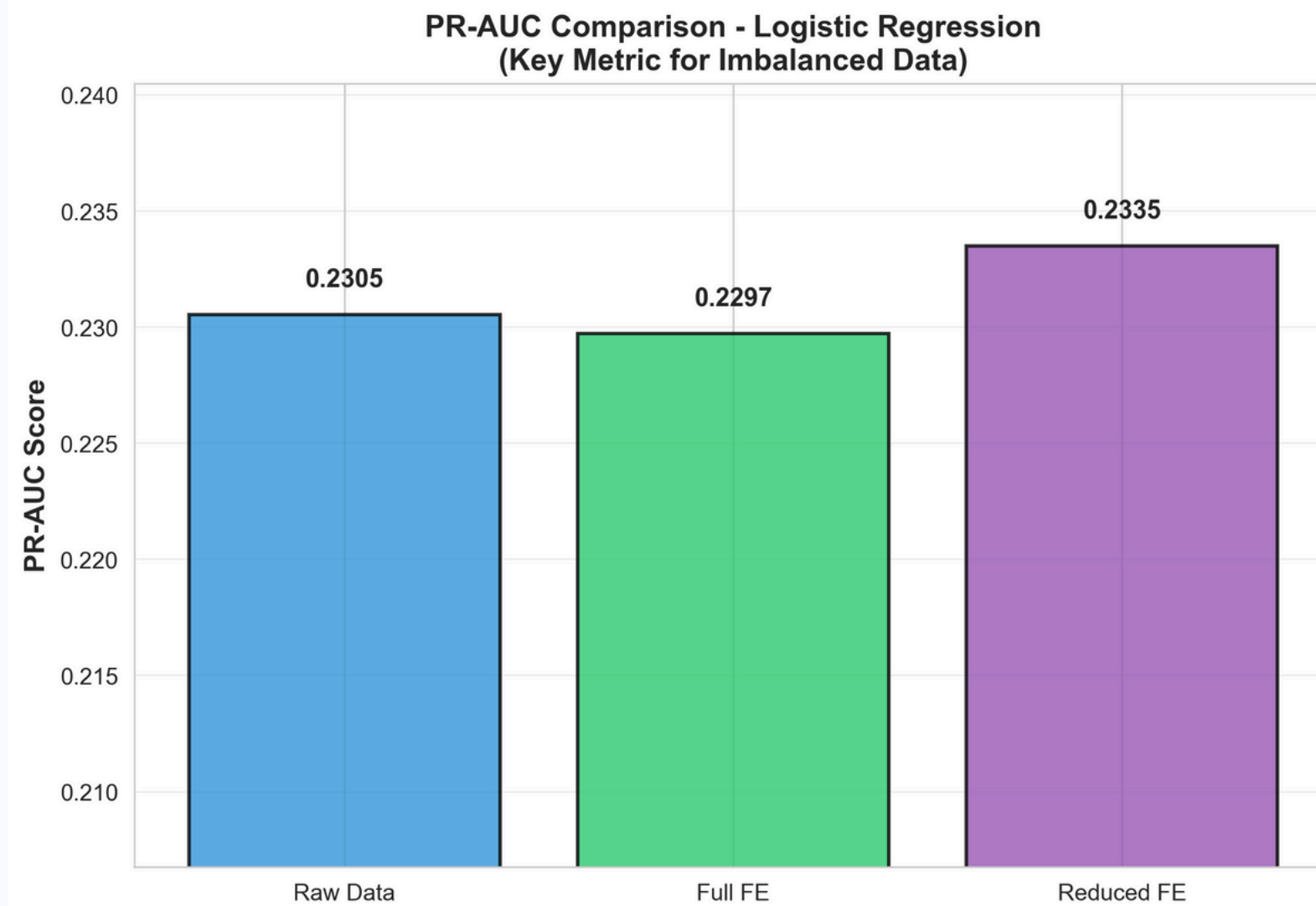
Al reducir de 46 a 32 features, el modelo pierde algo de capacidad discriminativa, aunque sigue siendo competitivo y más eficiente

# CONFUSION MATRICES LOGISTIC REGRESSION

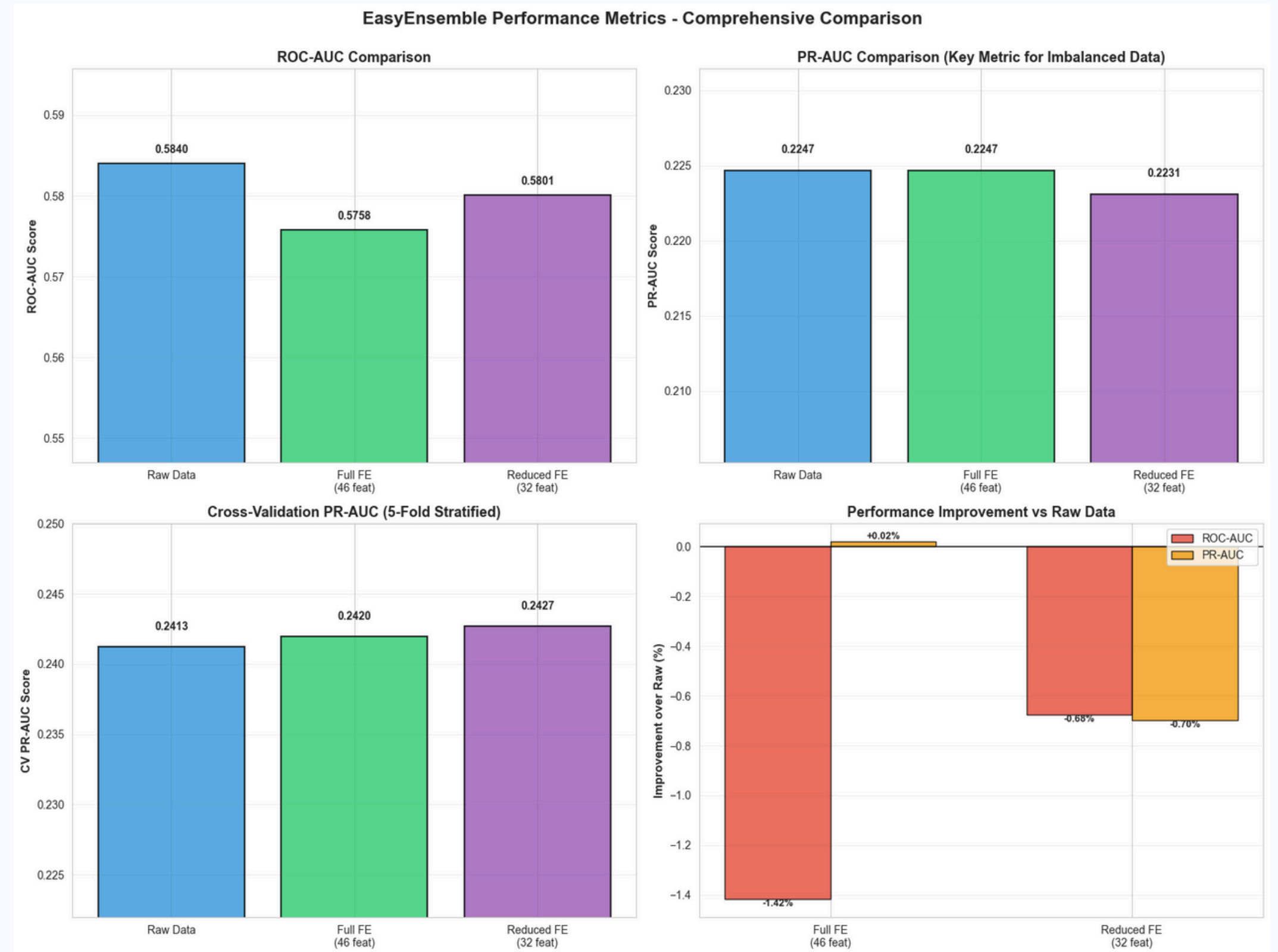


# PERFORMANCE LOGISTIC BARCHART

Logistic Regression Performance Metrics



# PERFORMANCE METRICS BAR CHART





**MUCHAS GRACIAS**