

2º Presentación Deep learning

**Integrantes: Alan Acuña
Felipe Mena
Profesor: Francisco Plaza**



Indice

- ▶ Objetivo General
- ▶ Objetivos especificos
- ▶ Descripción de datos y variables
- ▶ Estadística descriptiva variables usadas
- ▶ Distribucion de variables
- ▶ Diagramas de caja de variables
- ▶ Correlacion de variables
- ▶ Esquema metodologico
- ▶ Descripcion Modelo de deep learning 1
- ▶ Descripcion Modelo de deep learning 2



Objetivo General

Obtener el mejor modelo de clasificacion binaria en la detección de cuentas fraudulentas en el contexto de Criptomonedas de la plataforma de codigo abierto ETHEREUM

Hipotesis actual:

Evaluar si el modelo de redes neuronales de tres capas internas tiene un mejor desempeño frente al modelo logistico y SVM



Objetivos Especificos



1

Identificar las variables mas correlacionadas a la variable de deteccion de cuentas que realizaron transacciones fraudulentas

2

Realizar un analisis exploratorio estadistico de las variables respecto a las cuentas que se realizaron y no realizaron fraude

3

Ajustar a partir de estas variables los tres modelos de clasificacion binaria definidos con el mejor umbral

4

A partir de las metricas de precision comparar el desempeno de los tres modelos

Descripcion de datos



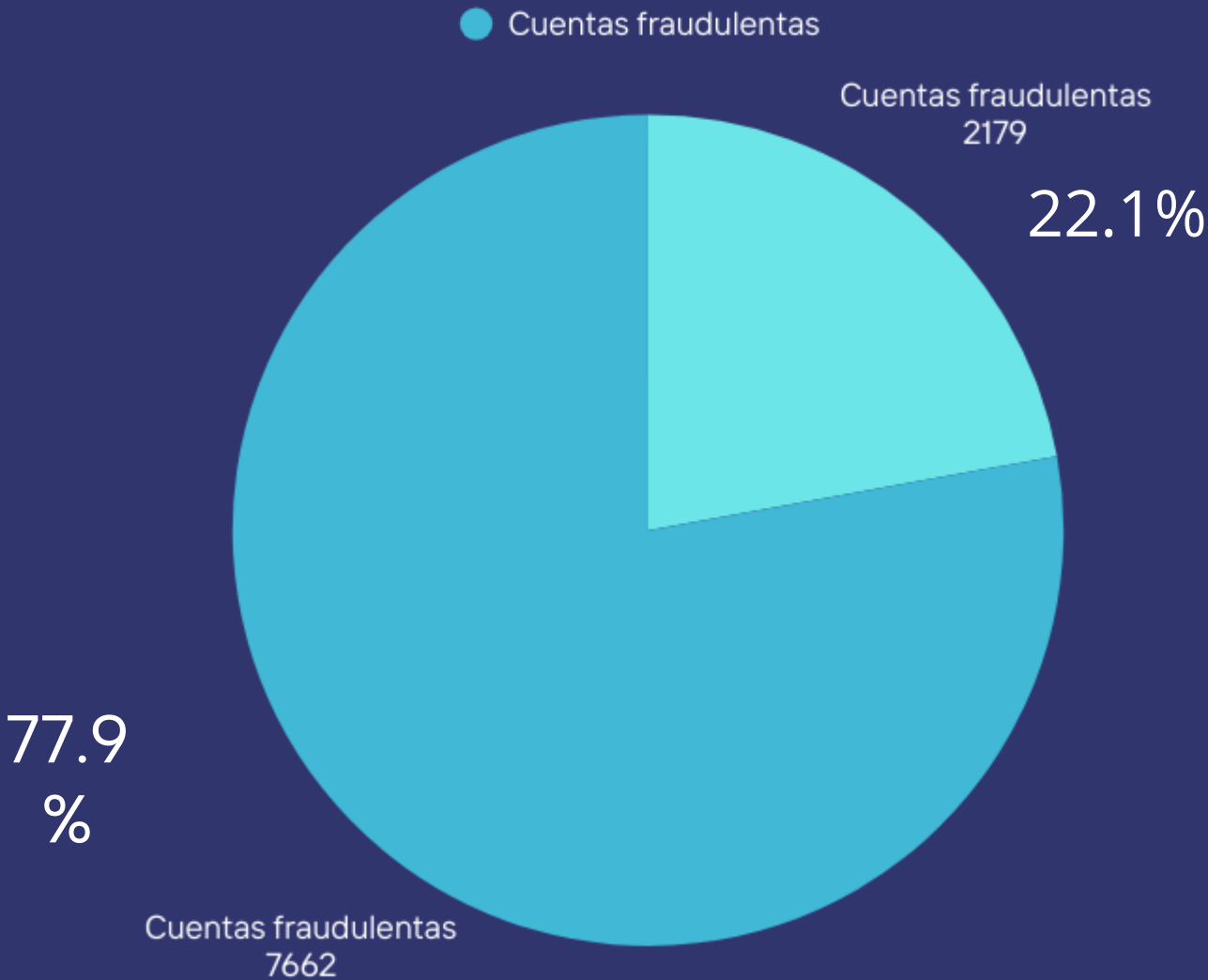
Base de datos

52 variables tipo numerico e informacion de 9841 cuentas fraudulentas



Variable de deteccion de cuenta Fraudulenta 'FLAG'

2179 cuentas fraudulentas 7662 cuentas no fraudulentas



Avg min between received tn timer

Tiempo promedio entre transacciones recibidas por cuenta en minutos.

Time Diff between first and last (Mins)

Diferencia horaria entre primera y la ultima transacción de la cuenta

Sent tn timer

Número total de transacciones normales enviadas

Received Tn timer

Número total de transacciones normales recibidas

total transactions

Numero total de transacciones de la cuenta

avg val sent

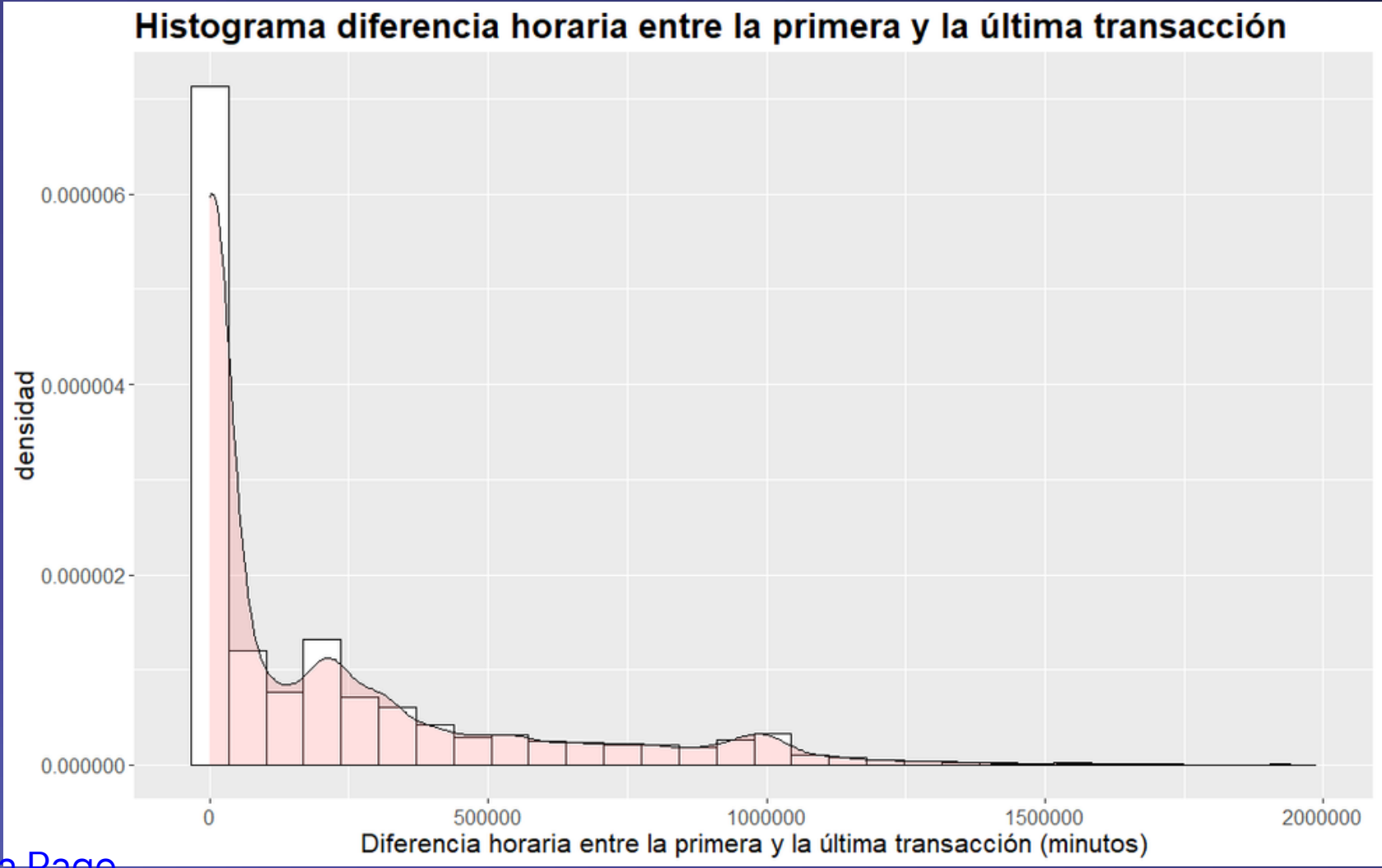
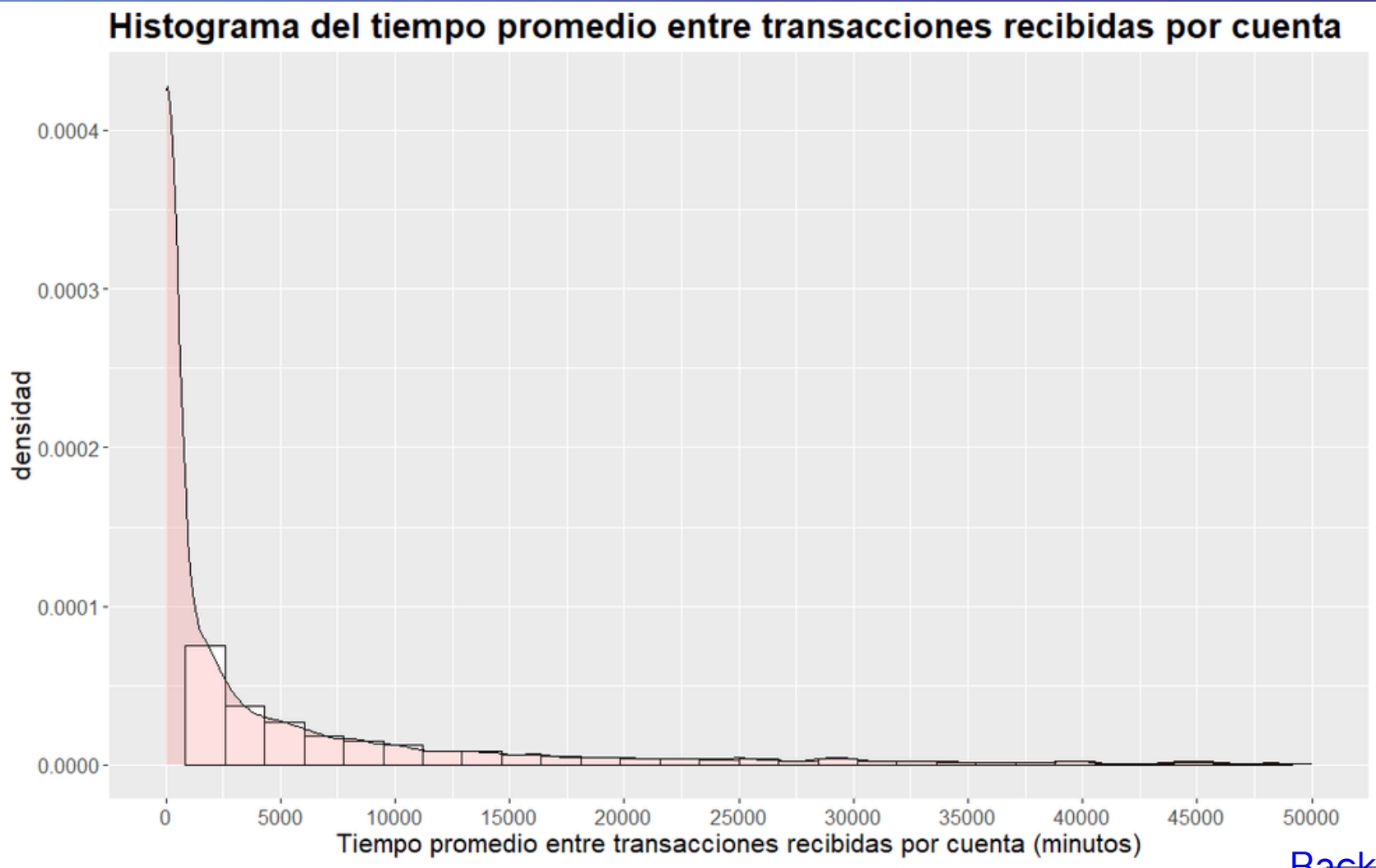
tiempo promedio entre transacciones enviadas en la cuenta en minutos

Variables	Correlaciones
avg val sent	-0.042854118
Received Tn timer	-0.062347083
Sent tn timer	-0.063147447
total transactions (including tn timer to create contract)	-0.079979555
Avg min between received tn timer	-0.115185945
Time Diff between first and last (Mins)	-0.212408063

Estadísticas descriptivas y distribuciones



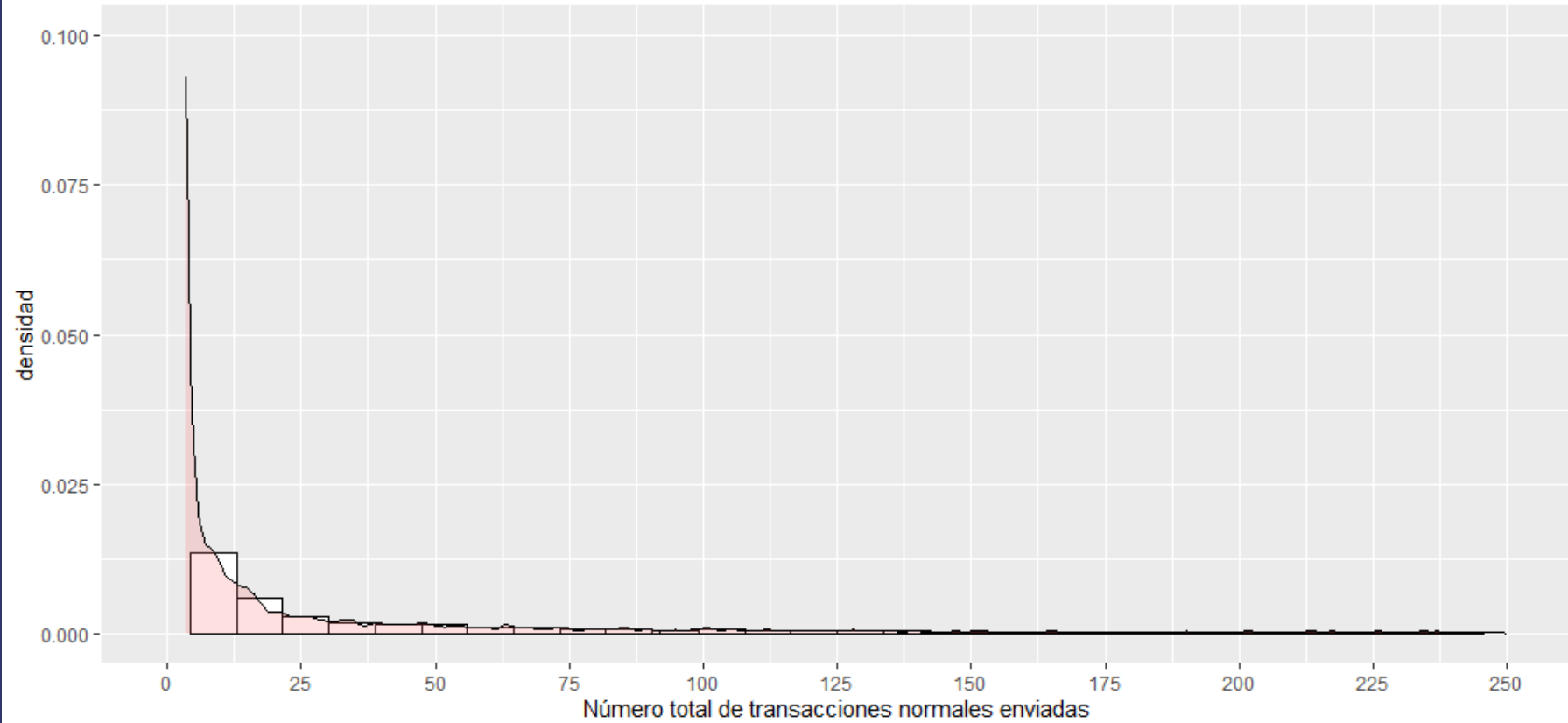
Variable	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
FLAG	0.0000	0.0000	0.0000	0.2214	0.0000	1.0000
Avg. min. between received txn	0.0	0.0	509.8	8004.9	5480.4	482175.5
Time Diff. between first and last (Mins)	0.0	316.9	46637.0	218333.3	304071.0	1954861.0
Sent txn	0.0	1.0	3.0	115.9	11.0	10000.0
Received txn	0.0	1.0	4.0	163.7	27.0	10000.0
Total transactions (including txn to create contract)	0.0	4.0	8.0	283.4	54.0	19995.0
Avg. val. sent	0.000	0.086	1.606	44.756	21.999	12000.000



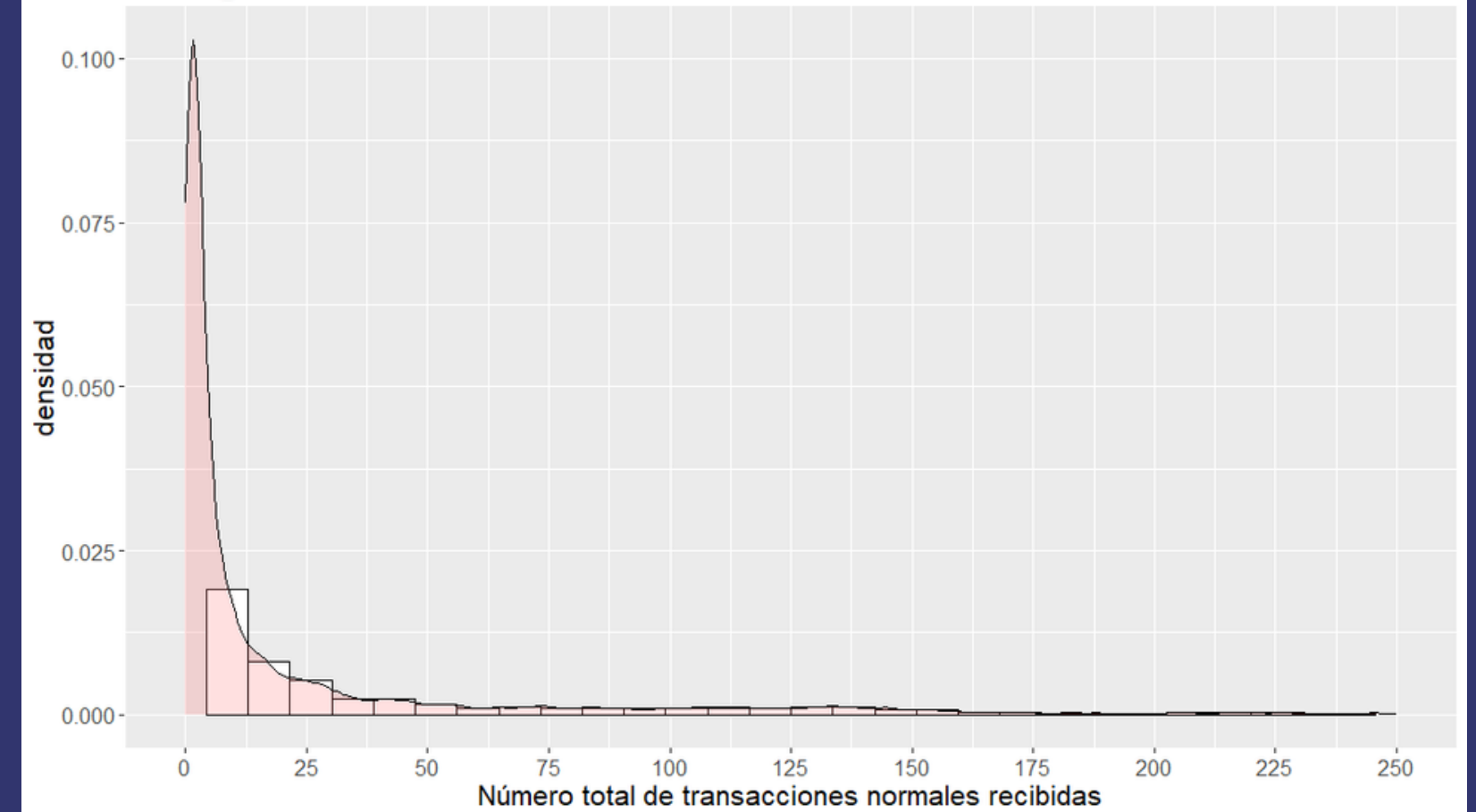
[Back to Agenda Page](#)

Distribuciones de las variables

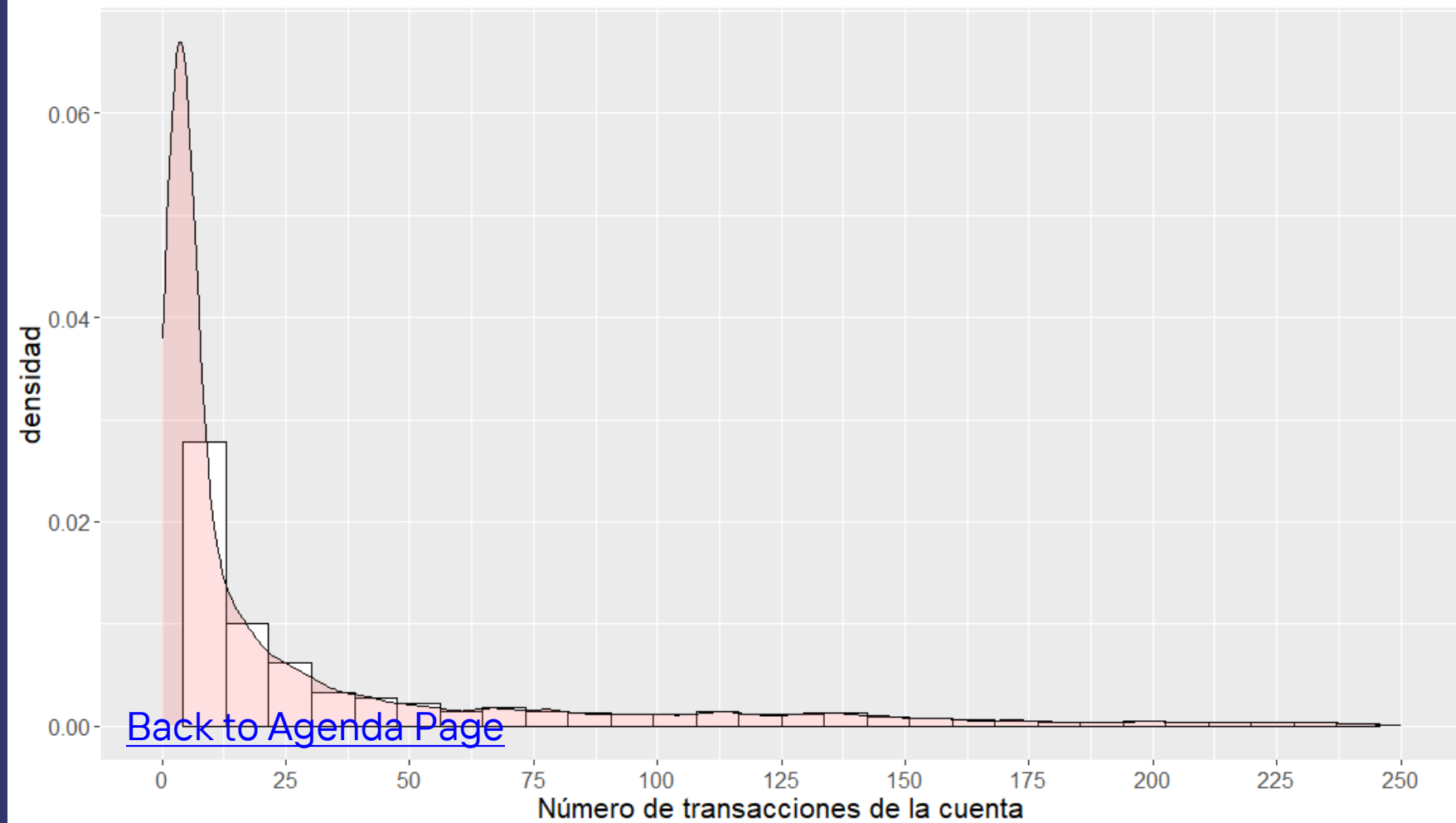
histograma Número total de transacciones normales enviadas



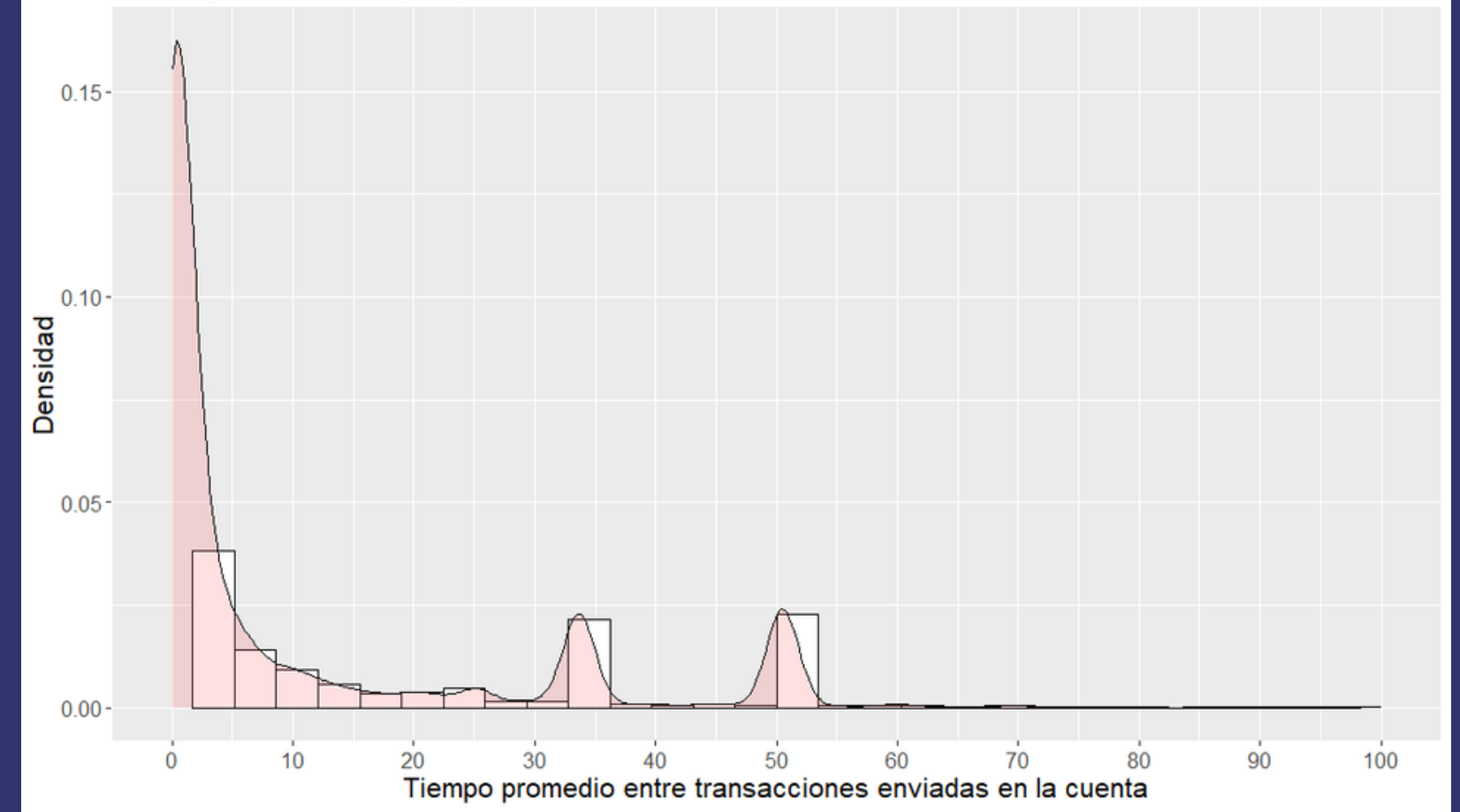
Histograma Número total de transacciones normales recibidas



Histograma Número de transacciones de la cuenta



Histograma tiempo promedio entre transacciones enviadas en la cuenta



[Back to Agenda Page](#)

Grafico de cajas

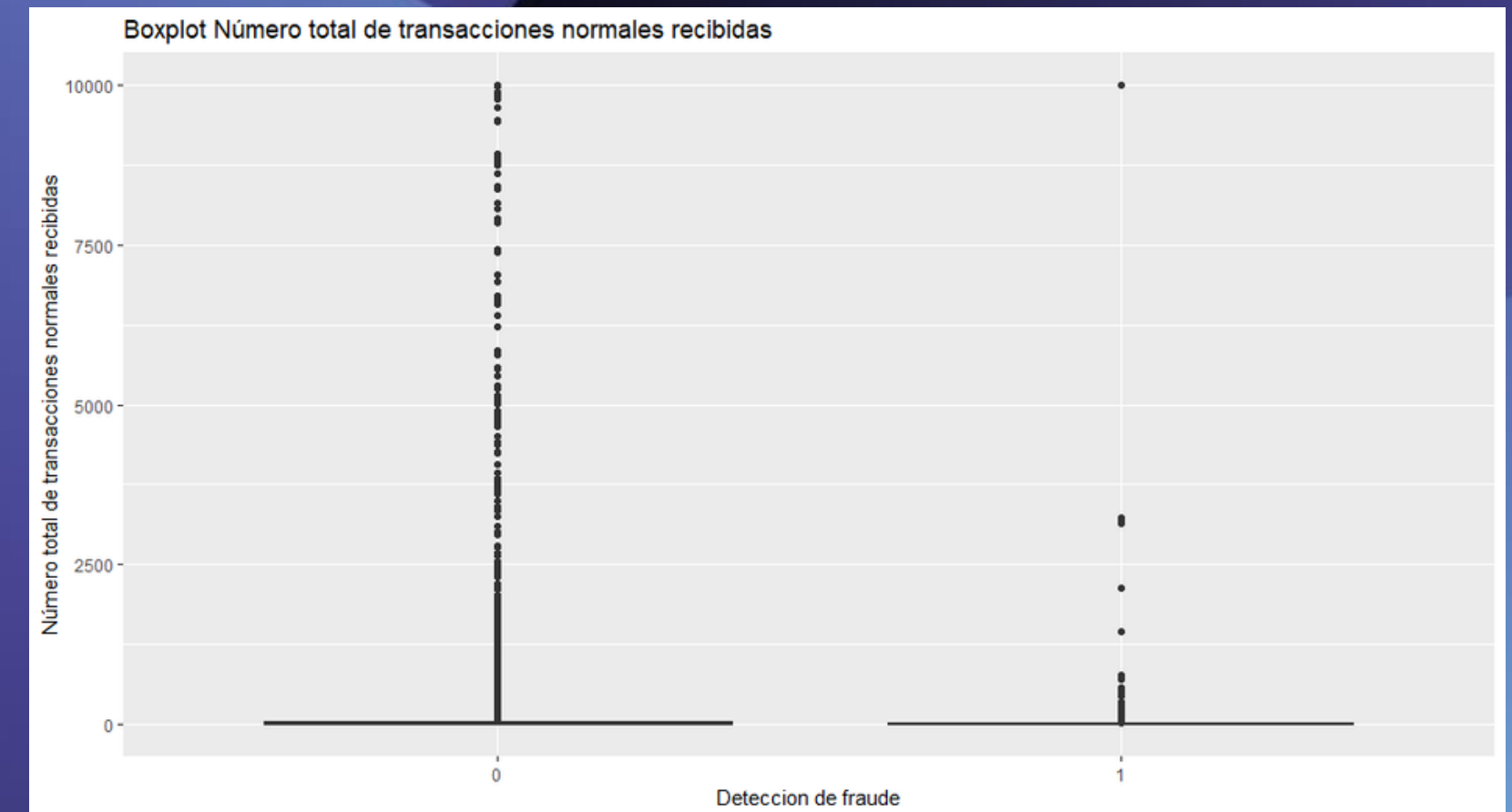
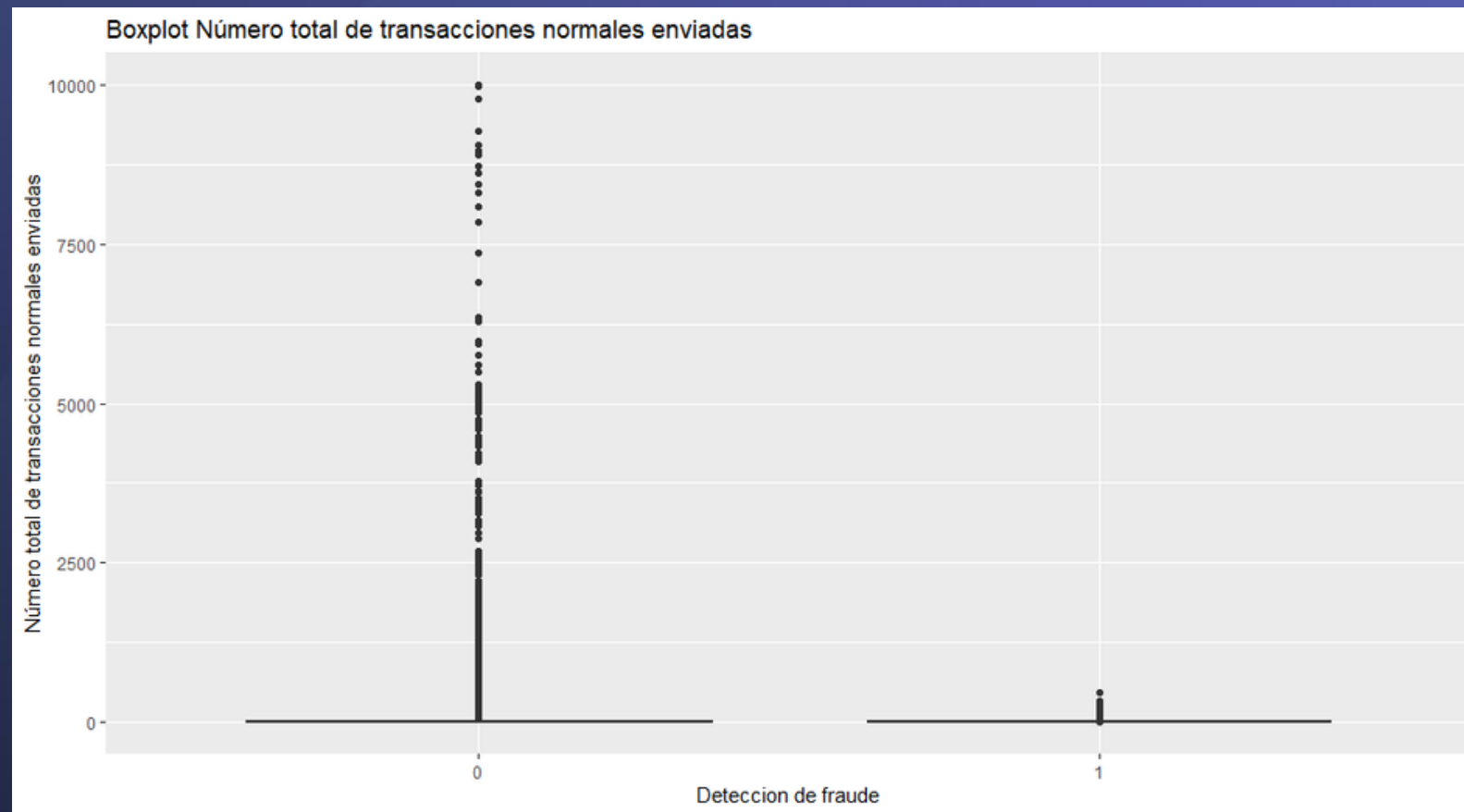
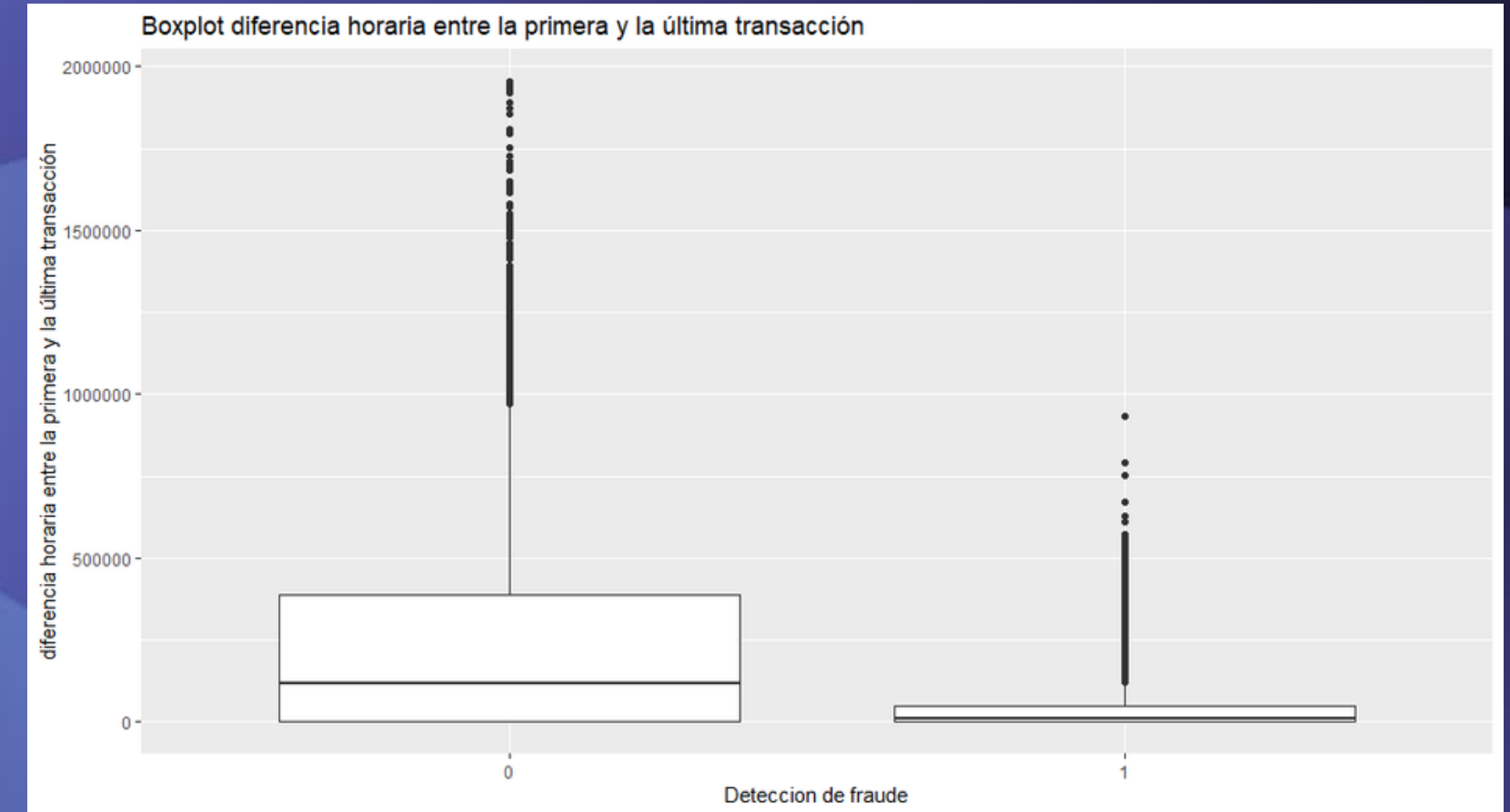
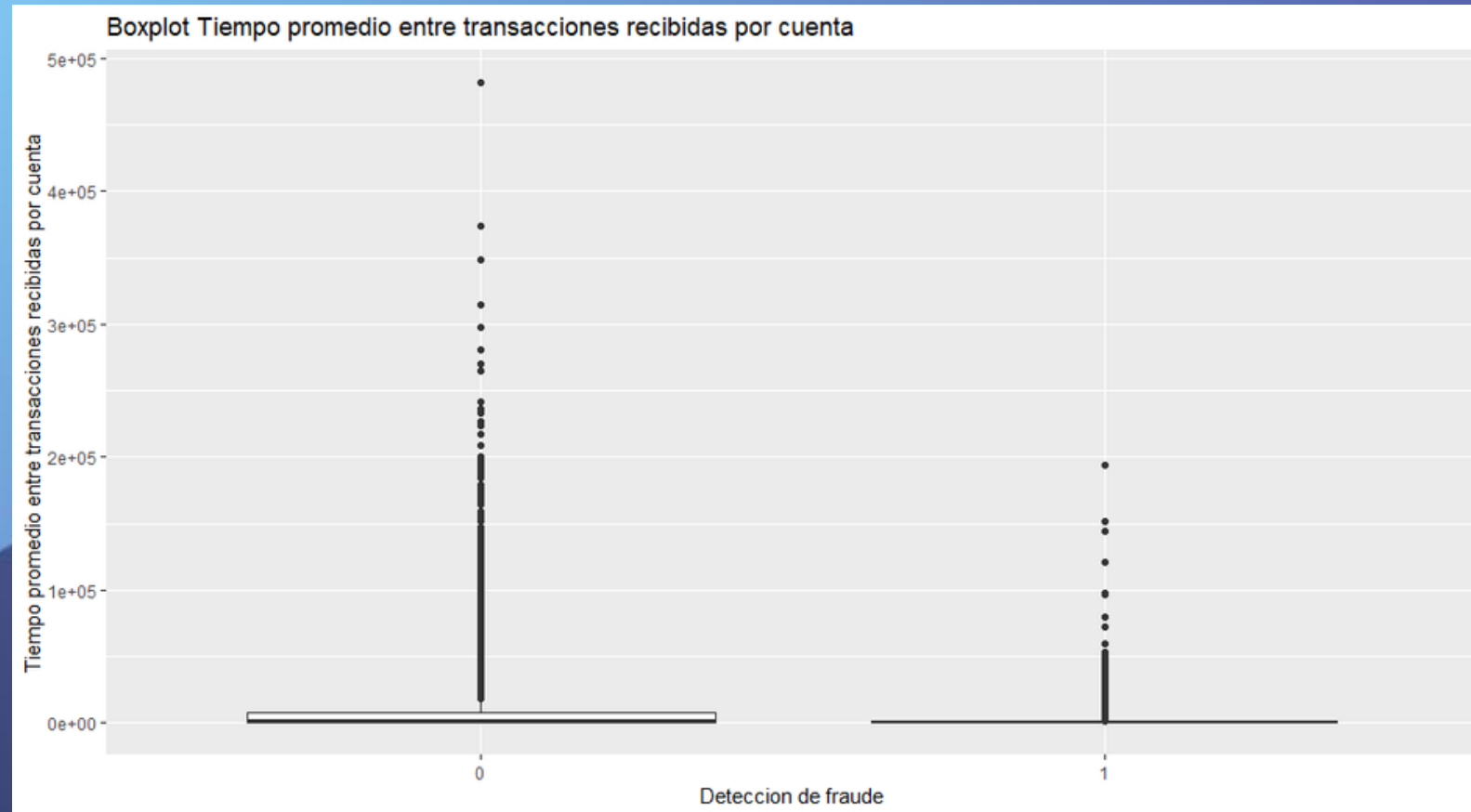
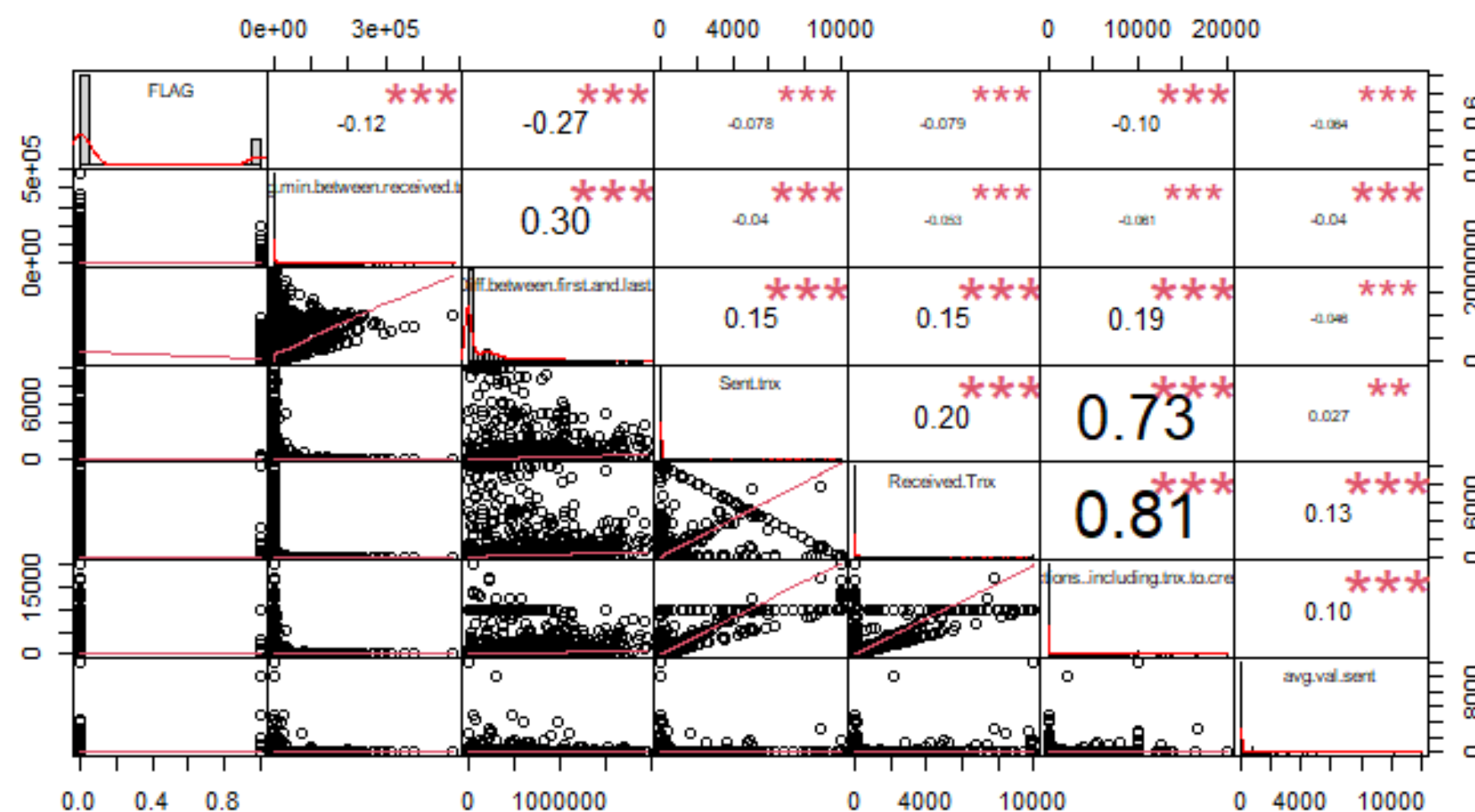
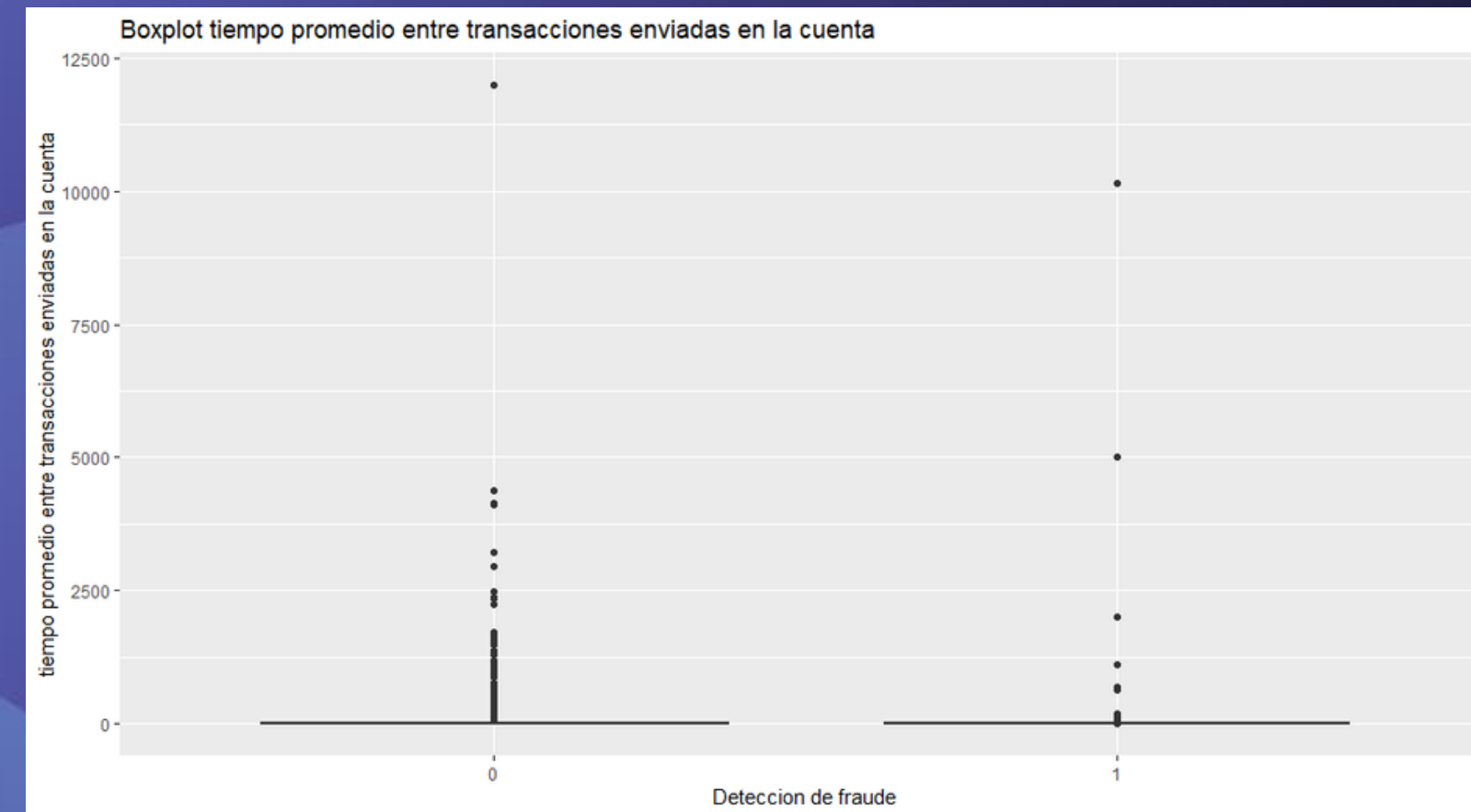
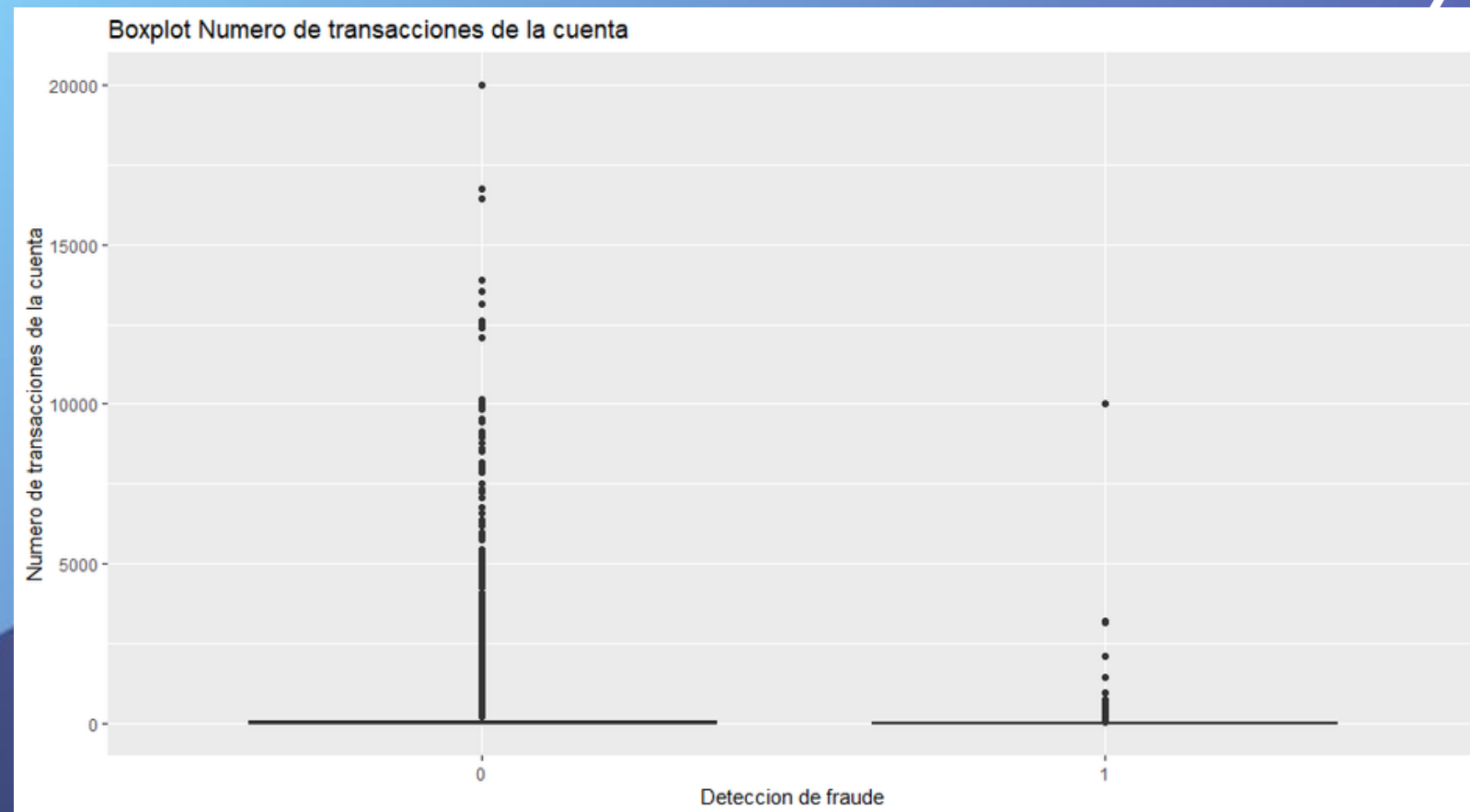


Grafico de cajas y correlaciones

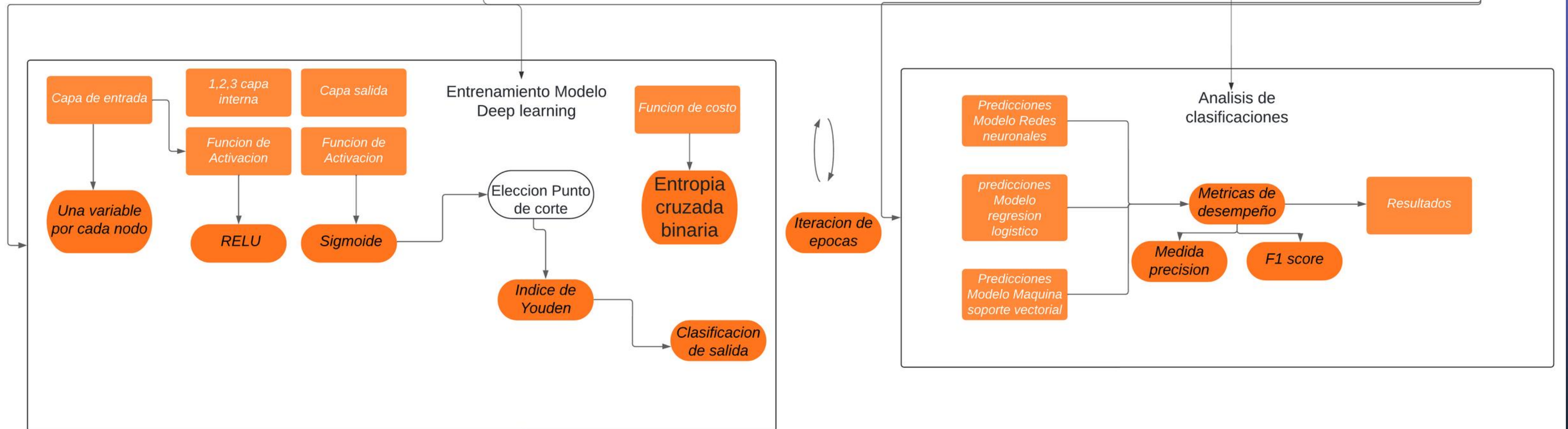
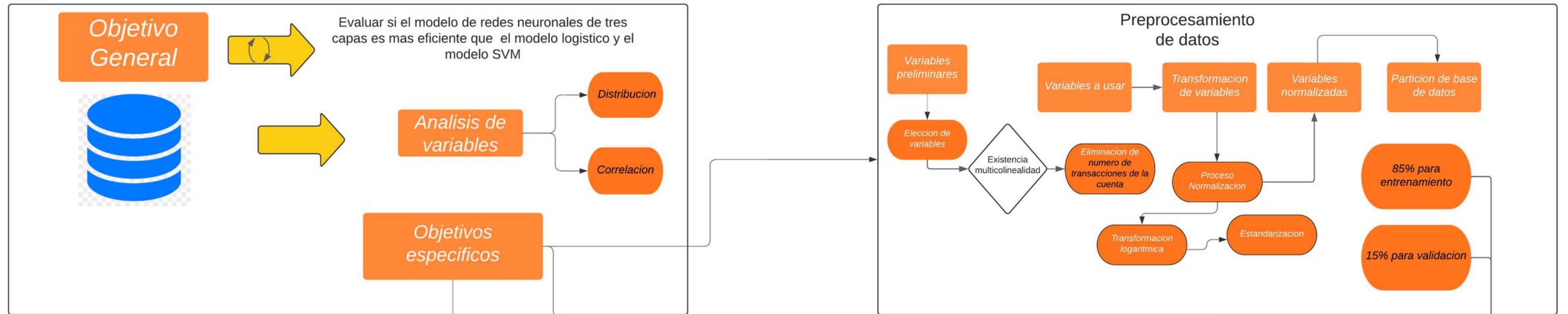


Problema de multicolinealidad

- total de transacciones
- número de transacciones normales enviadas
- número de transacciones normales recibidas

Esquema Metodológico general

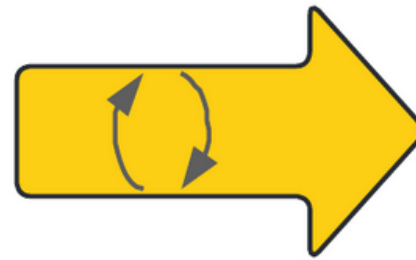
Analisis del problema



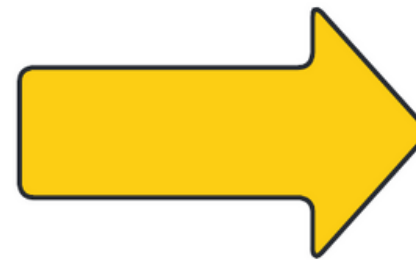
Esquema Metodológico parte 1

Analisis del problema

Objetivo General



Evaluar si el modelo de redes neuronales de tres capas es mas eficiente que el modelo logistico y el modelo SVM



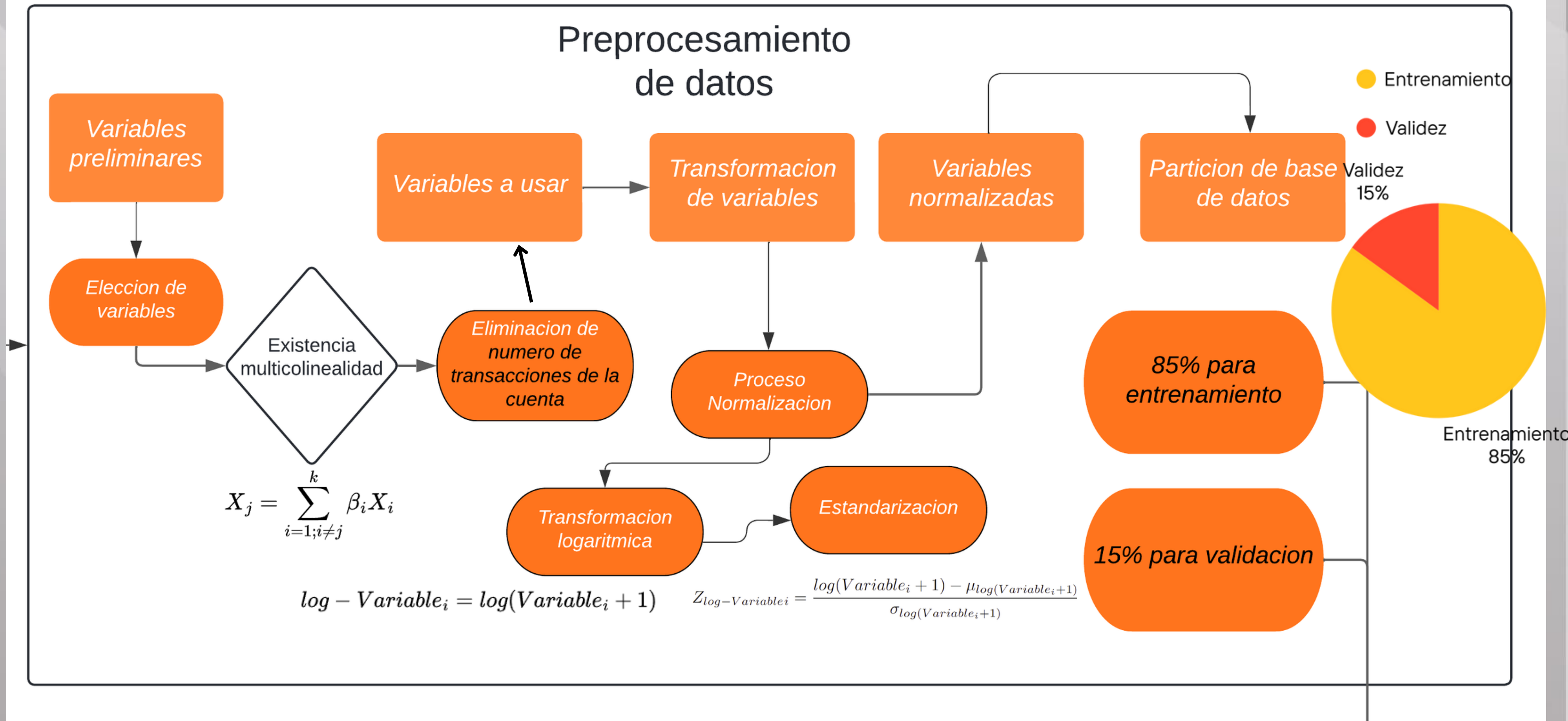
Analisis de variables

Distribucion

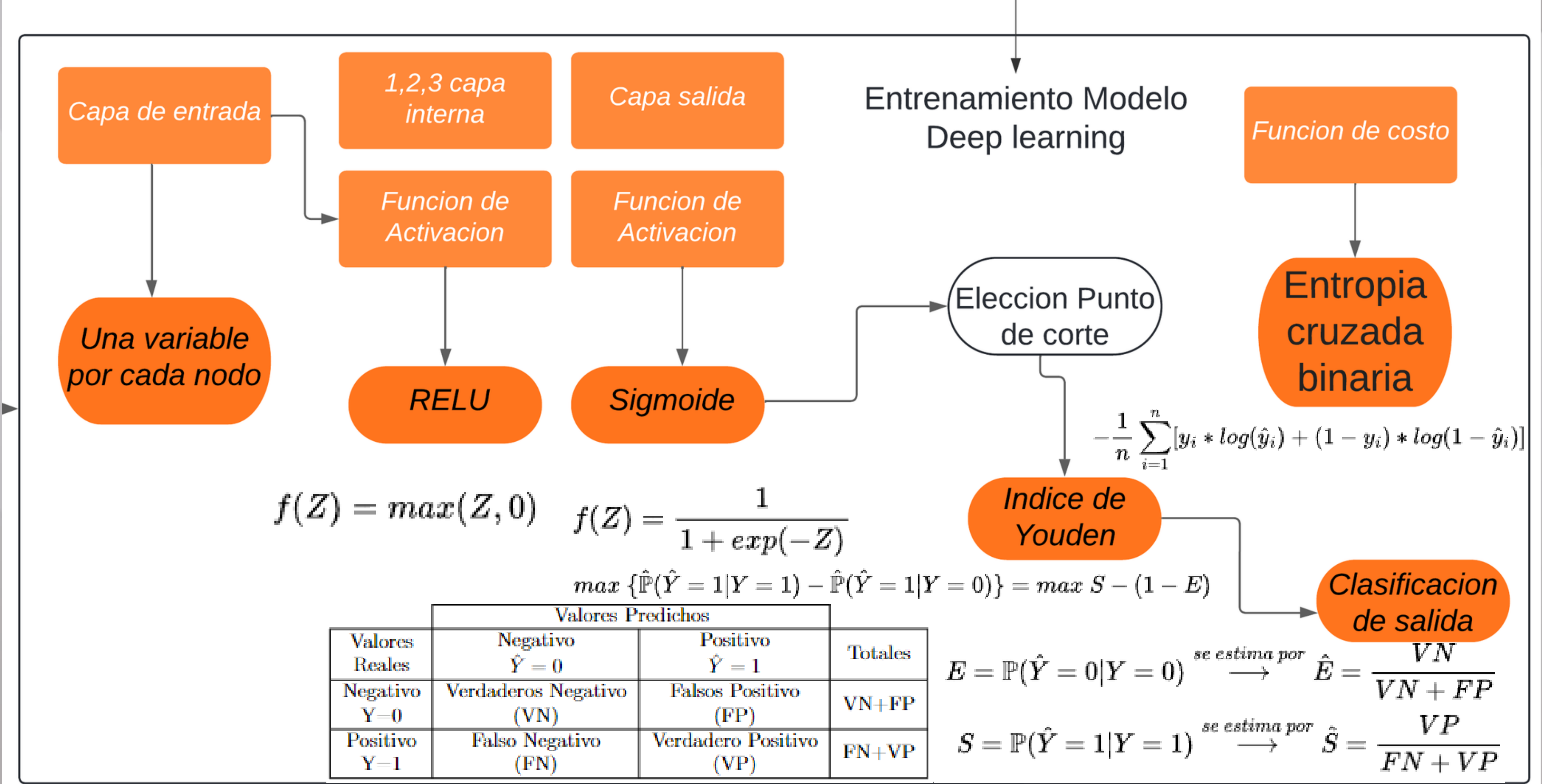
Correlacion

Objetivos especificos

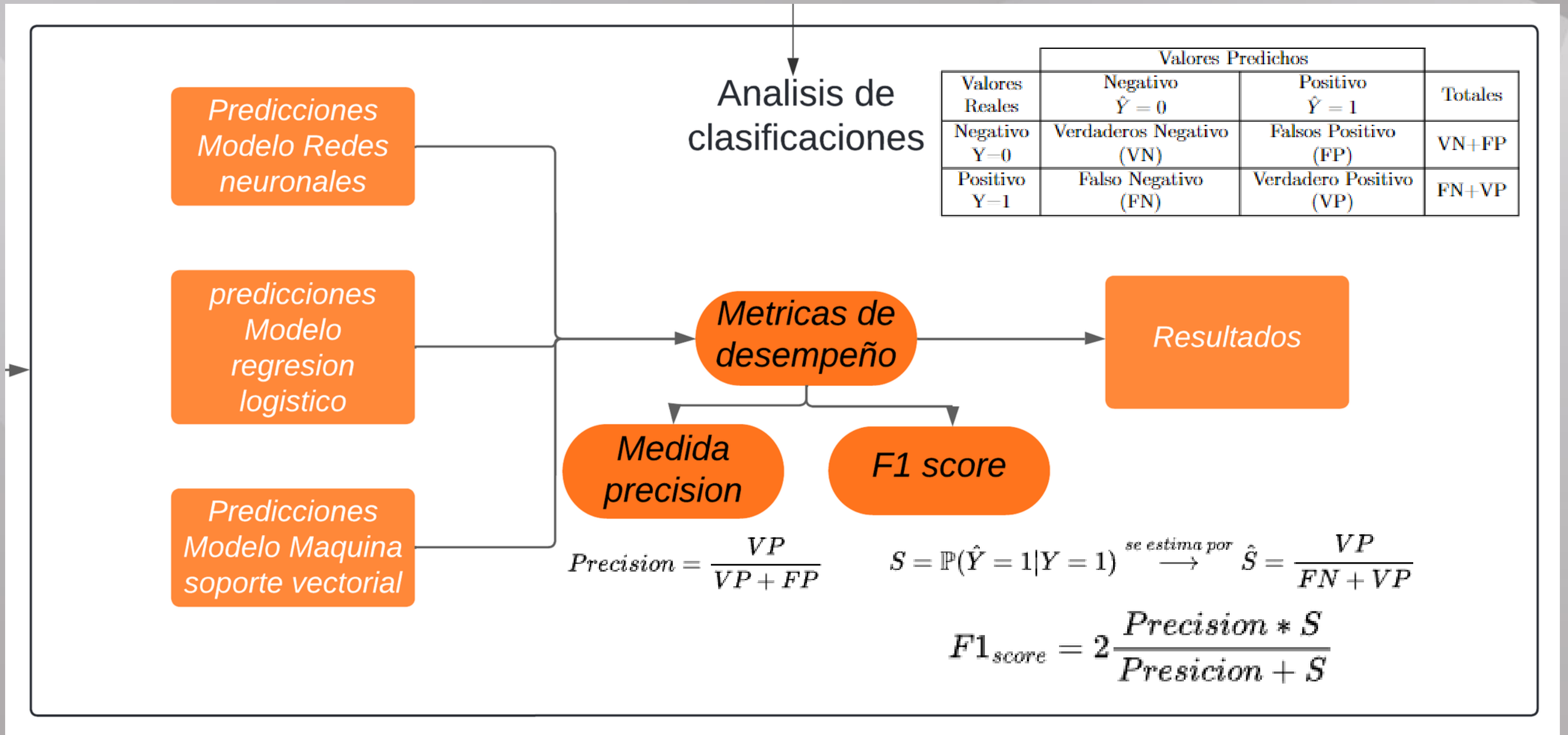
Esquema Metodológico parte 2



Esquema Metodológico parte 3



Esquema Metodológico parte 4



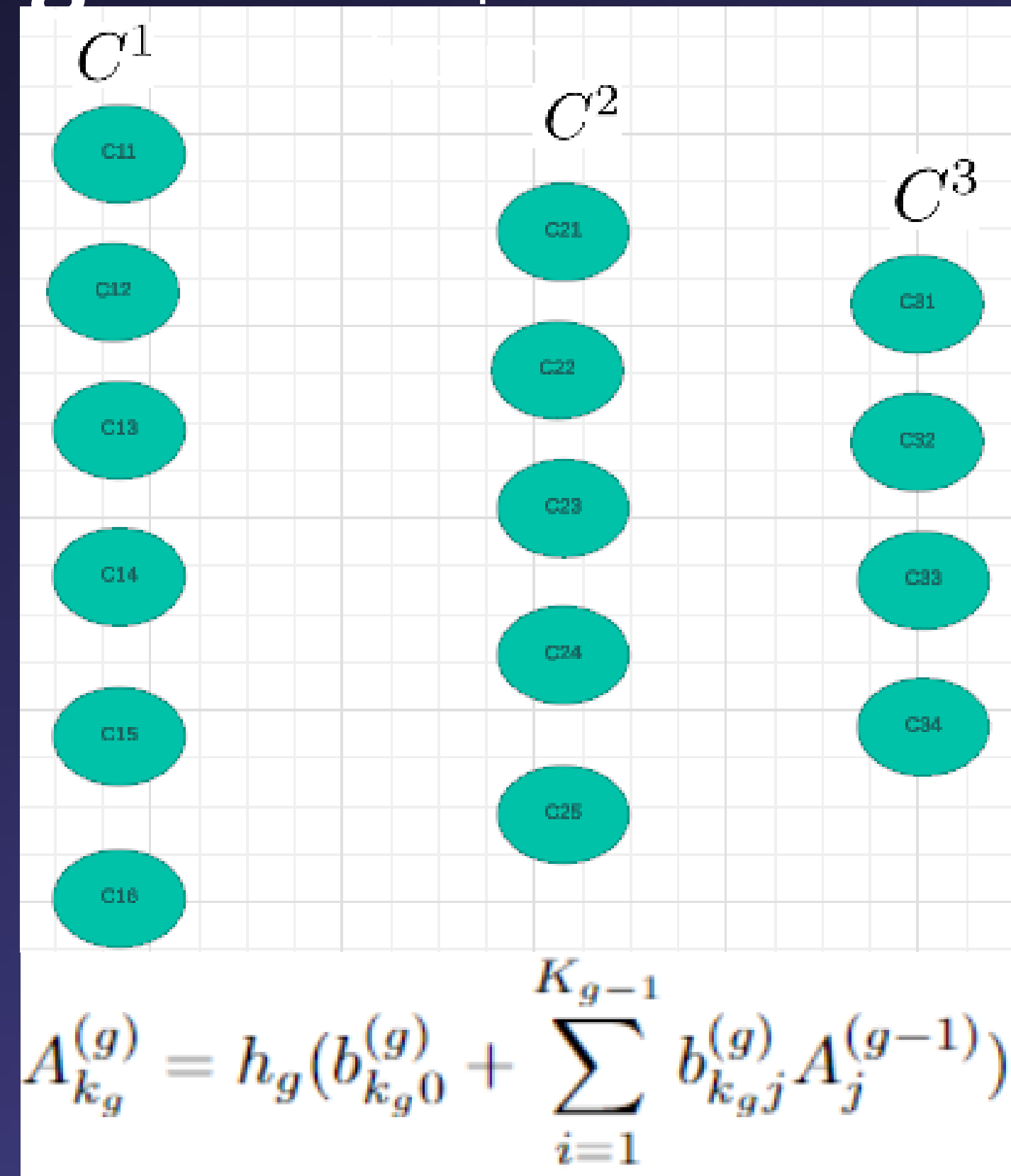
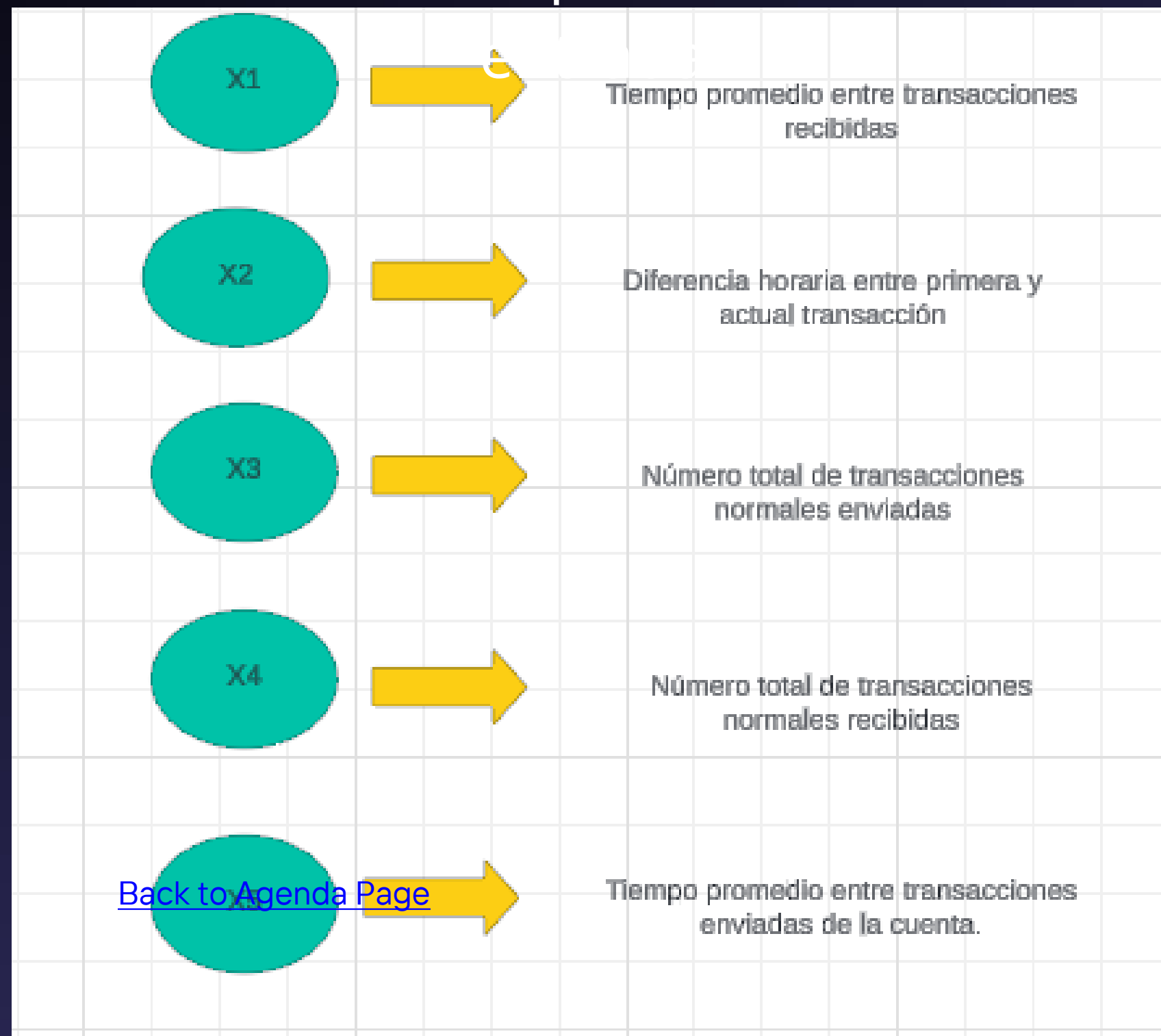
Descripción del modelo de Deep

Descripcion del modelo: Redes neuronal
de 3 capas internas

Learning

Capas

Capa de



$$h(a) = \max(0, a)$$

Descripción del modelo de Deep

Tasa de aprendizaje: Adam(Adaptive Moment Estimation)
Learning

Ultima Capa

T=0.001

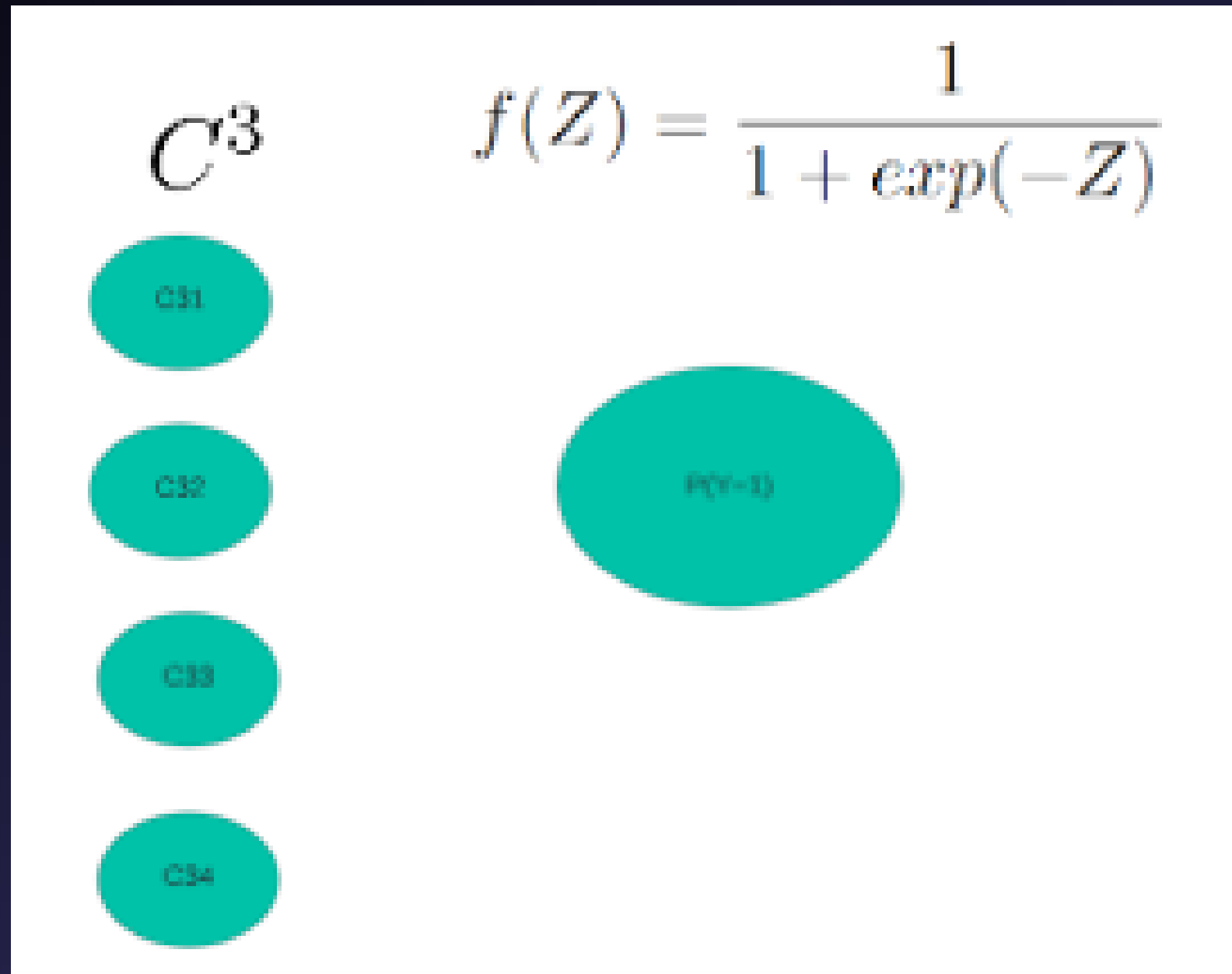
$$W_{t+1} = W_t - \alpha_t \frac{\tilde{u}}{\sqrt{\tilde{v} + \epsilon}}$$

Funcion de Perdida: Funcion entropía cruzada binaria

$$-\frac{1}{n} \sum_{i=1}^n [y_i * \log(\hat{y}_i) + (1 - y_i) * \log(1 - \hat{y}_i)]$$

Número de épocas : 15 épocas

Punto de corte para clasificar: Índice Youden



Valores Reales	Valores Predichos		Totales
	Negativo $\hat{Y} = 0$	Positivo $\hat{Y} = 1$	
Negativo $Y=0$	Verdaderos Negativo (VN)	Falsos Positivo (FP)	VN+FP
Positivo $Y=1$	Falso Negativo (FN)	Verdadero Positivo (VP)	FN+VP

$$\max \{ \hat{P}(\hat{Y} = 1|Y = 1) - \hat{P}(\hat{Y} = 1|Y = 0) \} = \max S - (1 - E)$$

$$E = \mathbb{P}(\hat{Y} = 0|Y = 0) \xrightarrow{\text{se estima por}} \hat{E} = \frac{VN}{VN + FP}$$

$$S = \mathbb{P}(\hat{Y} = 1|Y = 1) \xrightarrow{\text{se estima por}} \hat{S} = \frac{VP}{FN + VP}$$