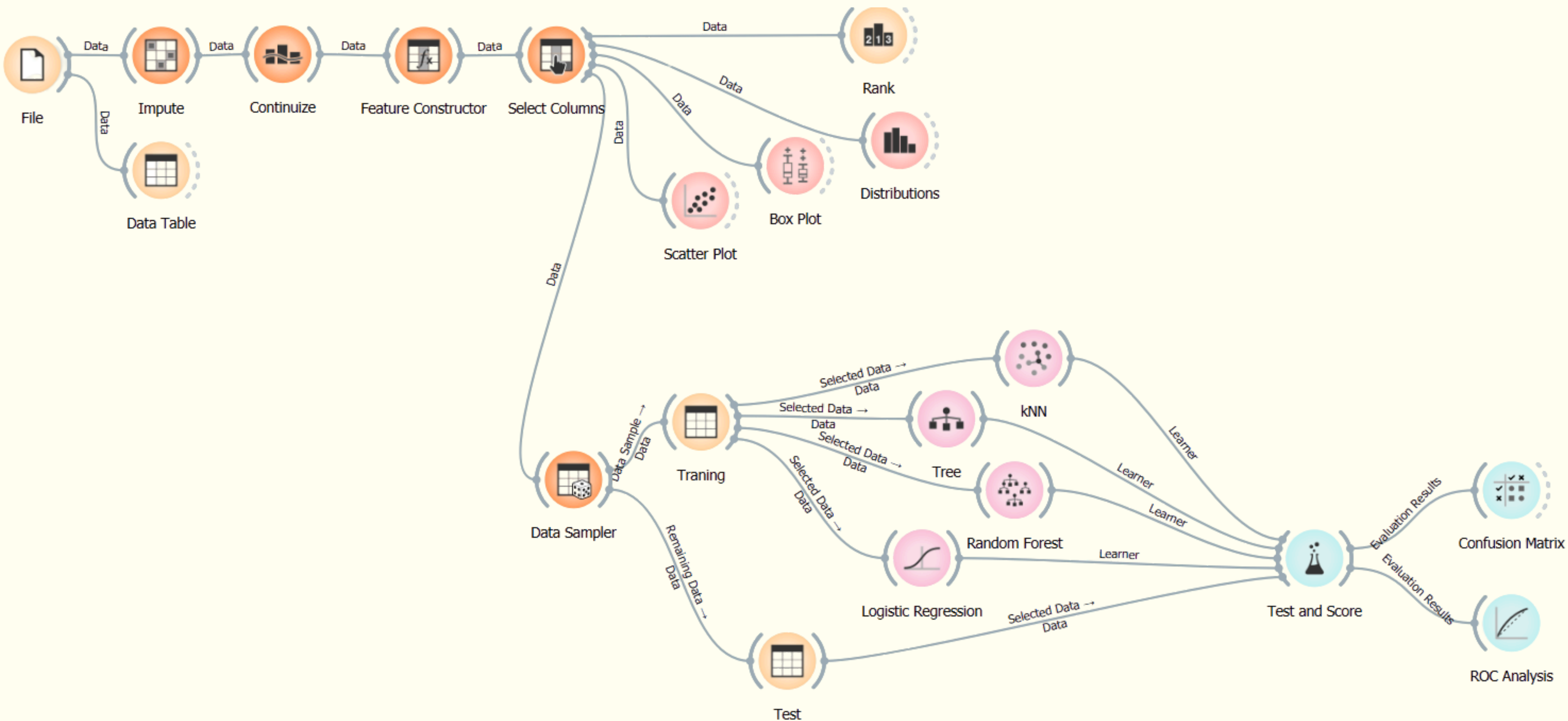




ORANGE

Marketing Campaign



Flujo ORANGE

Definición del Problema

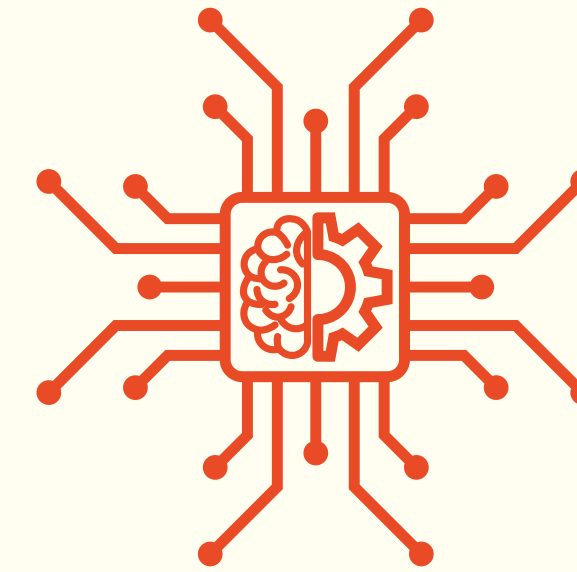
- ¿Qué buscamos predecir?

Si un cliente responderá positivamente a una campaña de marketing.

- ¿Por qué es importante?

Permite:

- Optimizar recursos de mkt
- Personalizar las campañas
- Aumentar la tasa de conversión

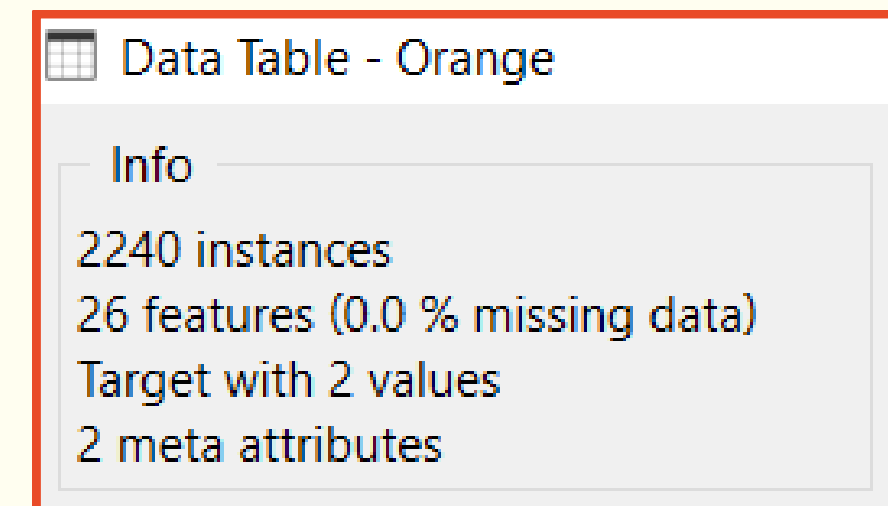


Objetivo del análisis

- Identificar los factores que influyen en la respuesta de los clientes y predecir quiénes tienen mayor probabilidad de aceptar una nueva oferta .

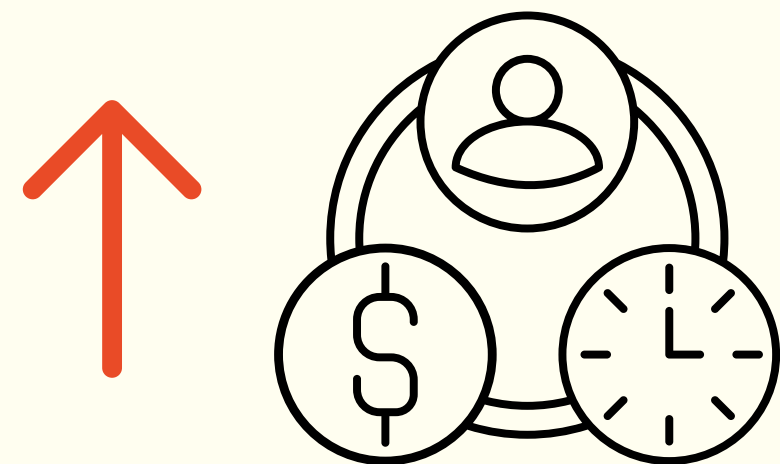
Introducción al Dataset

- Base de datos con 2.240 clientes de una campaña de marketing.
- Datos demográfica, económica y de consumo.
- Variable objetivo: Response (1=si, 0=no)





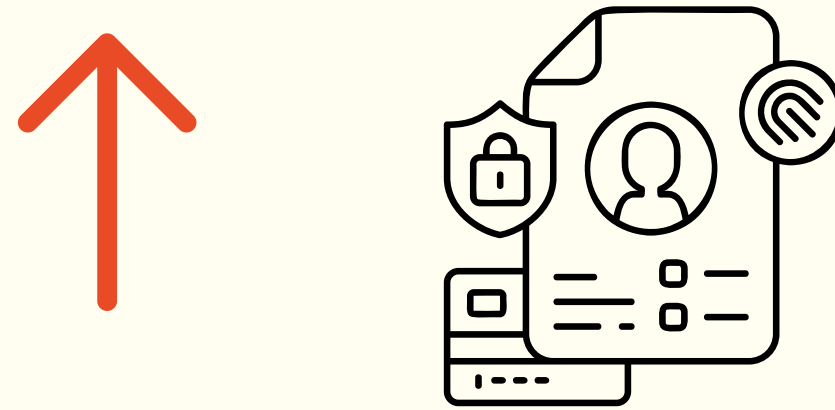
Variables principales



Días de la última compra

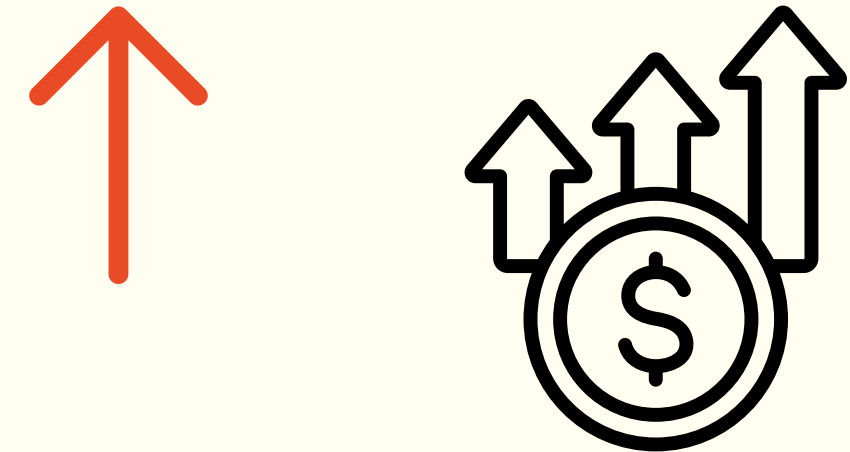
Recency

Datos personales



Edad, estado civil, nivel educativo.

Ingresos y Gastos



Nivel económico del cliente

TotalAccepted



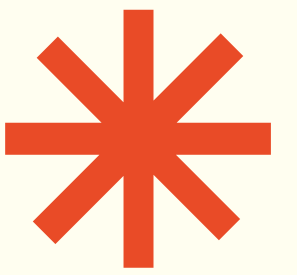
Cuántas campañas previas aceptó el cliente

Response



Variable objetivo
→ 1 = respondió, 0 = no respondió.

Hipótesis inicial del proyecto



Buscamos entender qué factores influyen en la aceptación de una campaña de mkt

1. Gasto total (Total_Spent)

Los clientes que gastan más tienen mayor probabilidad de aceptar campañas.

2. Ingreso (Income)

A mayor ingreso, mayor disposición a comprar.

3. Recency (última compra)

Los clientes recientes responden más.

4. Campañas previas (TotalAccepted)

Quienes ya aceptaron campañas son más propensos a hacerlo otra vez.

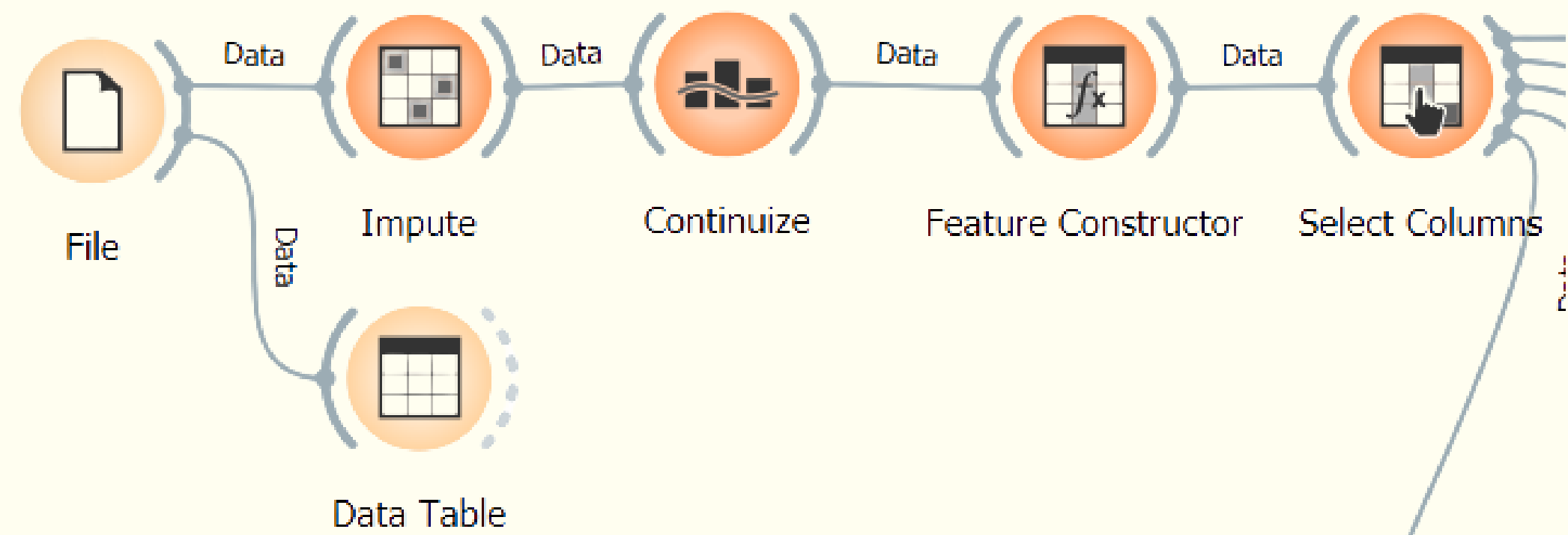
5. Educación (Education)

Puede influir, aunque en menor medida que las variables de gasto y comportamiento.

6. Correlación Ingreso ↔ Gasto

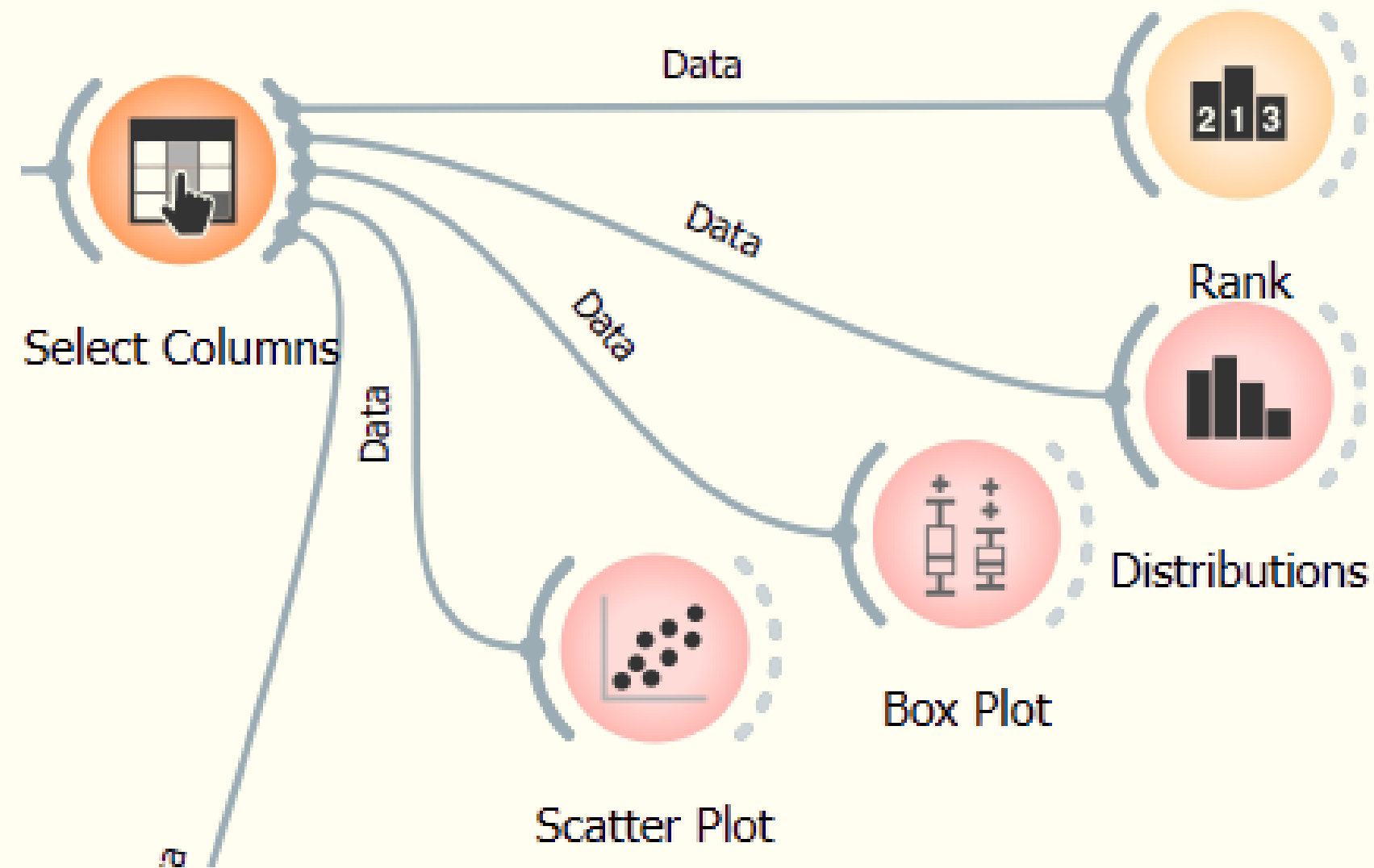
A mayor ingreso, mayor gasto total.

Preparación de datos



1. File	Importamos el dataset y definimos la variable objetivo (Response).
2. Impute	Completamos valores faltantes.
3. Continuize	Convertimos variables categóricas (Education, Marital_Status) a binarias (one-hot encoding).
4. Feature Constructor	Creamos nuevas variables: Age, TotalAccepted, TotalSpend.
5. Select Columns	Eliminamos las columnas que ya estaban agrupadas o no aportaban valor.

EDA- Análisis Exploratorio de datos



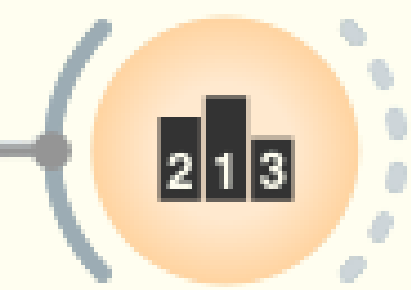
Objetivo:

Explorar patrones de comportamiento y factores que influyen en la respuesta a campañas

Principales hallazgos:

- **Mayor gasto (Total_Spent)** y mayores ingresos (Income) → más respuestas
- **Menor recency** → más participación.
- Variables más influyentes: **TotalAccepted, Total_Spent, NumCatalogPurchases, Recency, Income**

El cliente que responde suele ser activo, con alto gasto e ingresos medios-altos, y con historial de interacción positiva con la marca.



Importancia de Variables (Rank)

Rank Muestra el ranking de las variables con mayor influencia para predecir la respuesta (Response).

		#	Info. gain	Gain ratio	ANOVA
1	TotalAccepted		0.099	0.101	496.290
2	Total_Spent		0.045	0.022	169.443
3	NumCatalogPurchases		0.044	0.022	114.712
4	Recency		0.028	0.014	91.738
5	Income		0.024	0.012	40.151
6	NumWebPurchases		0.021	0.011	50.626
7	Teenhome		0.019	0.017	54.689
8	NumWebVisitsMonth		0.008	0.004	0.036
9	Marital_Status=Single		0.007	0.010	25.027
10	NumStorePurchases		0.007	0.004	3.473

Principales insights

TotalAccepted → mayor historial de aceptación = más chances de responder.

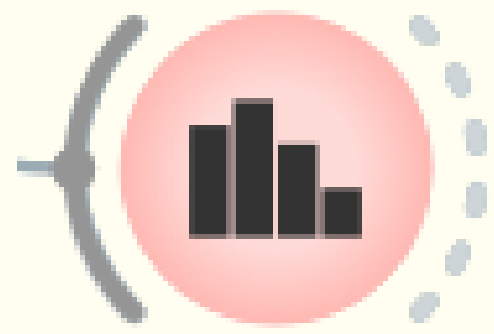
Total_Spent → mayor gasto total = mayor probabilidad de aceptación.

NumCatalogPurchases → quienes compran por catálogo responden mejor.

Recency → clientes recientes participan más.

Income → ingresos altos influyen positivamente.

Las variables de comportamiento previo y nivel de gasto son las que más explican la respuesta del cliente.



Income

Distributions

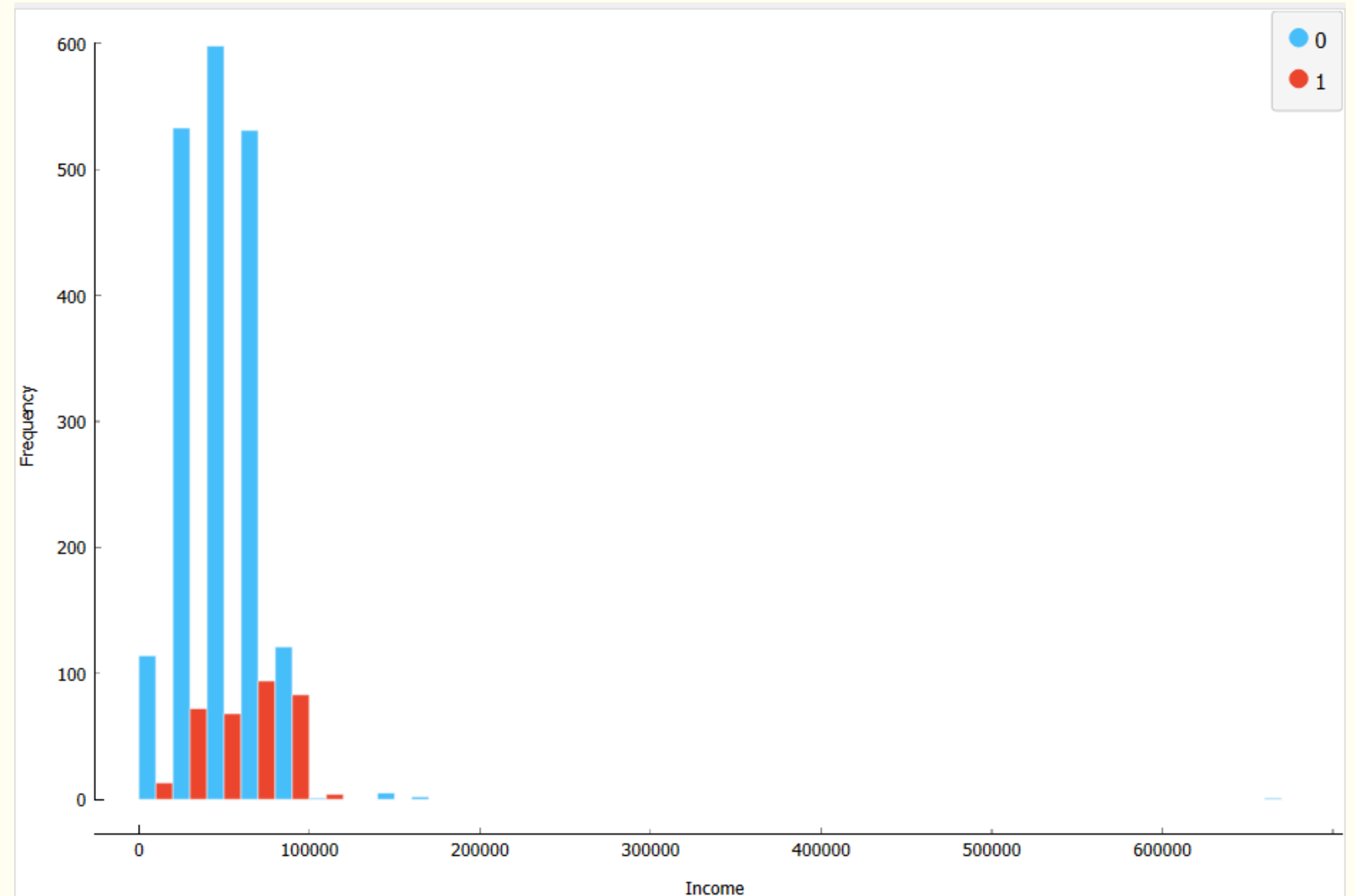
Ingreso en relacion a campaña aceptada

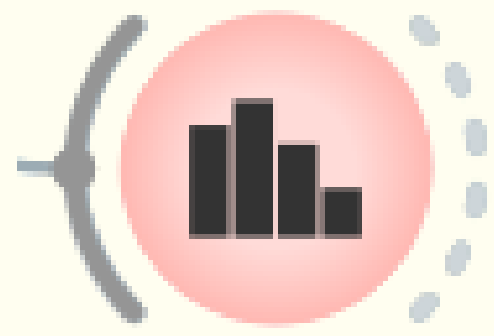
Rojo = Acepto

Azul = No Acepto

Observaciones:

- Mayor concentración de clientes con ingresos entre 30.000 y 100.000.
- Las respuestas positivas se dan sobre todo en niveles de ingreso medio-alto.
- Los clientes con más ingresos tienden a aceptar más campañas, aunque son menos en cantidad.





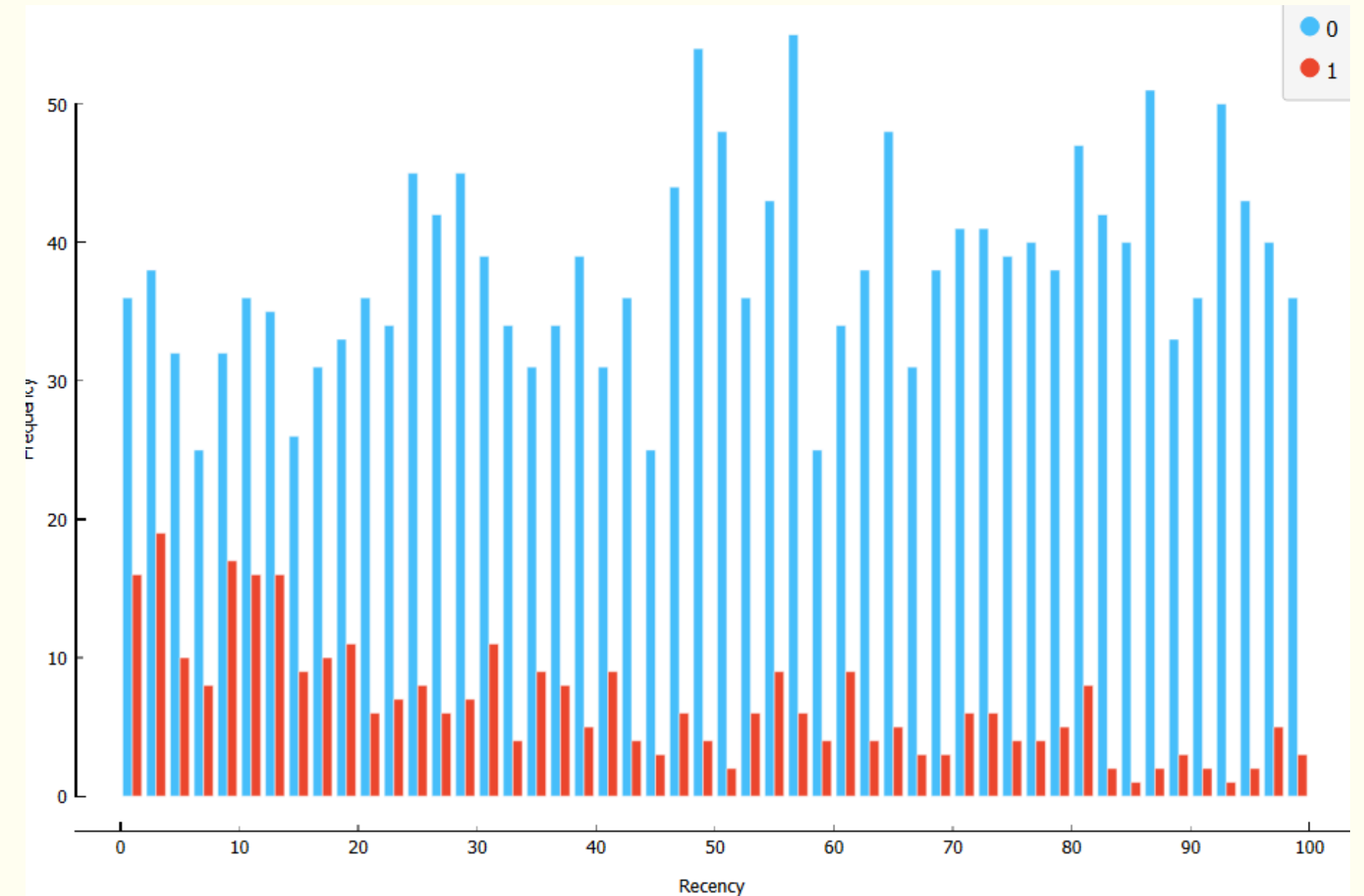
Recency

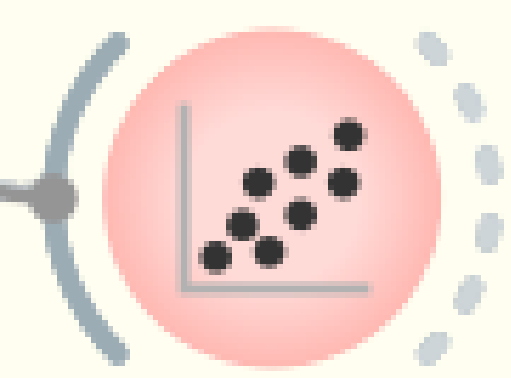
Distributions

El gráfico muestra la cantidad de clientes según los días transcurridos desde su última compra (Recency), diferenciando entre quienes aceptaron (1) y no aceptaron (0) la campaña.

Observaciones:

- Se observa que la mayoría de los clientes con alta recencia (más días sin comprar) no respondieron positivamente.
- Y que las respuestas afirmativas se concentran en valores bajos de recencia, es decir, en clientes que realizaron compras más recientes.





Ingreso - Gasto

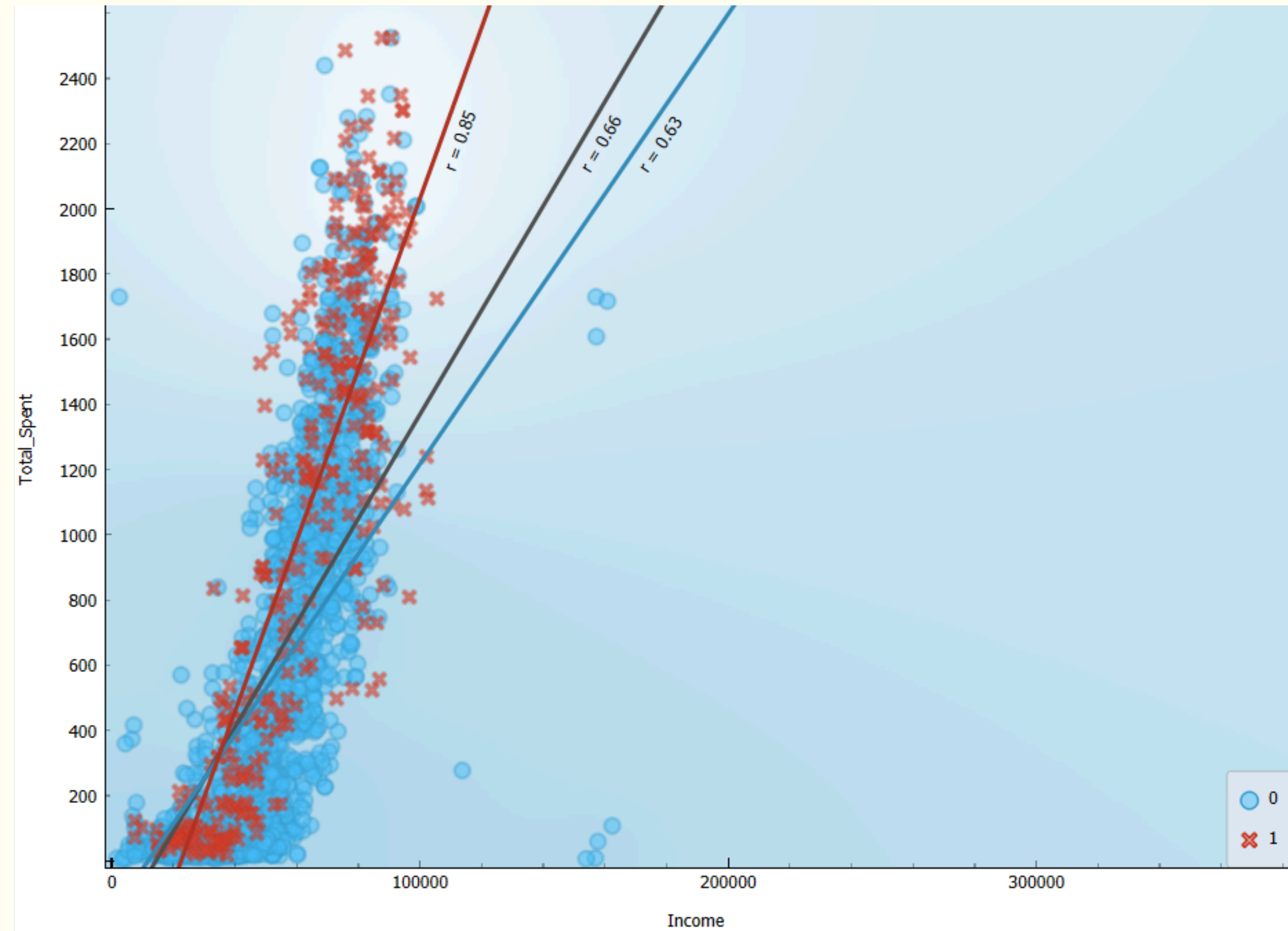
Scatter Plot

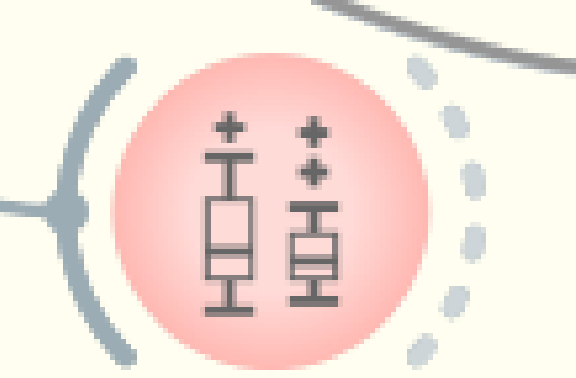
Relación entre el ingreso anual (Income) y el gasto total (Total_Spent) de los clientes.

- **Rojo** → Aceptaron
- **Azul** → No Aceptaron

Observaciones:

- Correlación positiva entre ingreso y gasto ($r = 0.67$ general), lo que indica que a mayor ingreso, mayor nivel de gasto.
- Clientes que aceptaron la campaña presentan una correlación más fuerte ($r = 0.85$), esto sugiere que los clientes de mayor poder adquisitivo y con hábitos de consumo altos son más propensos a responder positivamente.





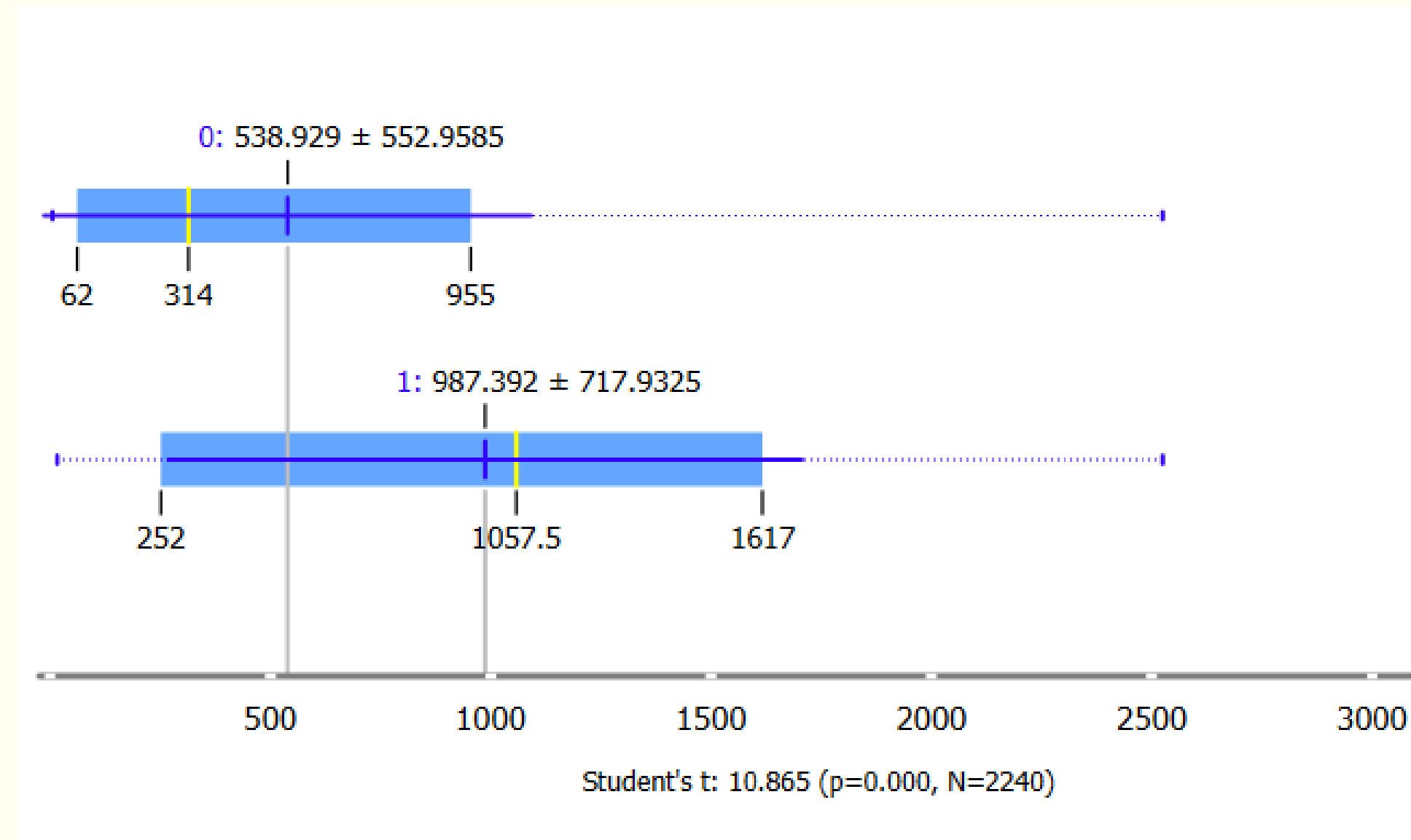
Gasto - Aceptación

Box Plot

El gráfico compara el gasto total (Total_Spent) entre los clientes que aceptaron y los que no aceptaron la campaña.

Observaciones

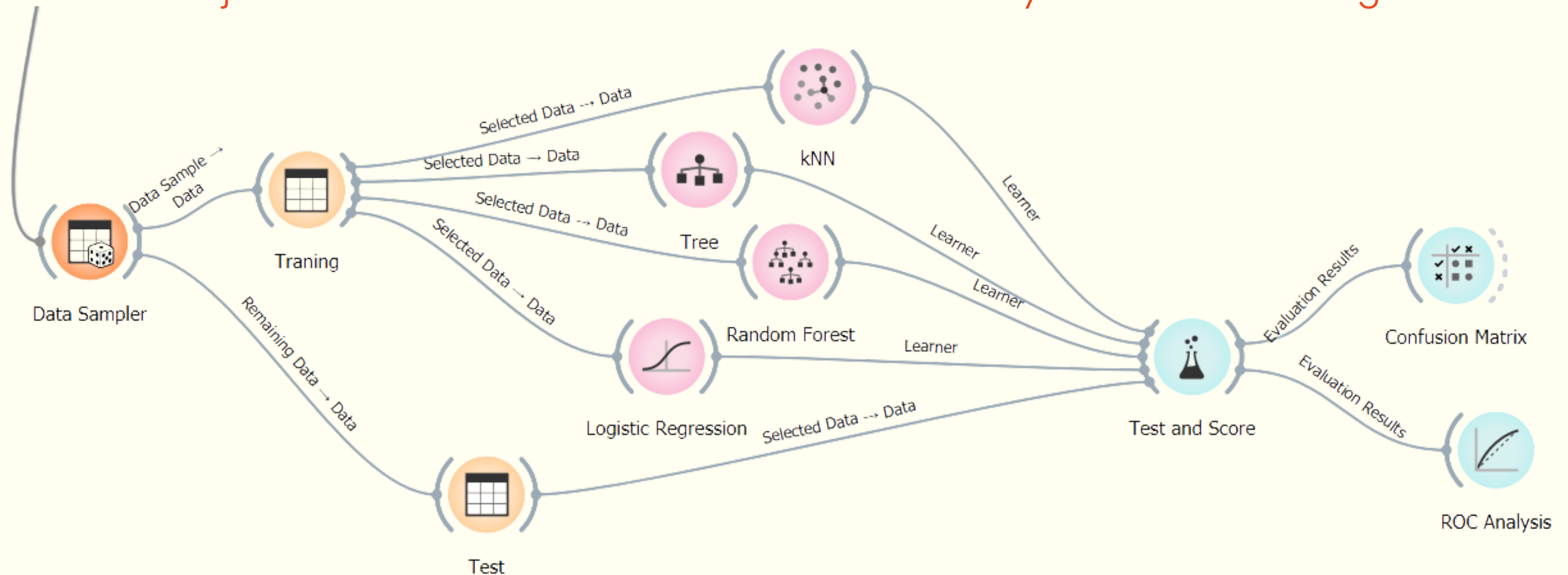
- Los clientes que aceptaron la campaña (1) presentan un gasto promedio mucho mayor (~987) que los que no aceptaron (0) (~539).
- Esto indica una relación positiva entre el nivel de gasto y la probabilidad de aceptación de una campaña.
- El valor $p = 0.000$ confirma que la diferencia entre ambos grupos es estadísticamente significativa.



En términos de negocio, los clientes con mayor gasto total son más propensos a responder positivamente, por lo que deberían ser prioritarios en futuras campañas de marketing.

División del Dataset (80% – 20%)

Objetivo: evaluar el rendimiento real del modelo y evitar el overfitting.

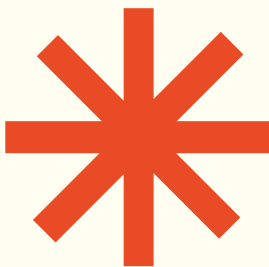


- 80% entrenamiento: el modelo aprende los patrones y relaciones entre variables.
- 20% prueba: se usa con datos nuevos no vistos para medir su desempeño real.
- Esta división permite validar la capacidad de generalización del modelo.
- Así garantizamos que no “memorice” los datos, sino que prediga correctamente casos nuevos.



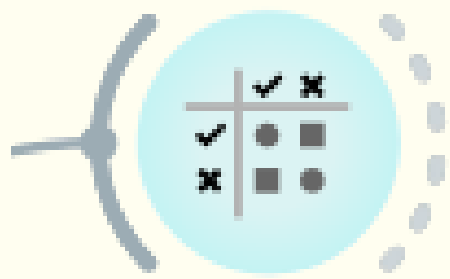
Resultados Test and score

Test and Score



Modelo	AUC	Accuracy	F1	Precision	Recall	Observaciones
Random Forest	0.866	0.873	0.846	0.861	0.873	Mejor balance global entre precisión y recall. Captura bien ambas clases.
Logistic Regression	0.821	0.868	0.846	0.85	0.868	Buen rendimiento, ligeramente inferior al Random Forest.
kNN	0.651	0.844	0.805	0.804	0.844	Menor capacidad predictiva, sensible al ruido.
Tree	0.633	0.846	0.834	0.828	0.846	Sencillo pero menos preciso, tendencia a sobreajustar.





Random forest

Matriz de confusión



Confusion Matrix

Verdadero Negativo – 372 casos

- El modelo predijo que el cliente no aceptó la campaña y la realidad era que no aceptó.

Falso Positivo – 6 casos

- El modelo predijo que el cliente aceptó la campaña y la realidad era que no aceptó.

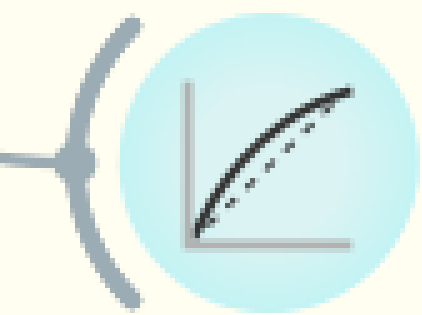
Falso Negativo – 51 casos.

- El modelo predijo que el cliente no aceptó la campaña y la realidad era que sí aceptó.

Verdadero Positivo – 19 casos

- El modelo predijo que el cliente aceptó la campaña y la realidad era que sí aceptó.

		Predicted		Σ
		0	1	
Actual	0	372	6	378
	1	51	19	70
Σ		423	25	448



Análisis de desempeño

Curva ROC

ROC Analysis

Muestra la capacidad de los modelos para distinguir entre clases, y para predecir correctamente si un cliente aceptara (1) o no aceptara (0) la campaña.

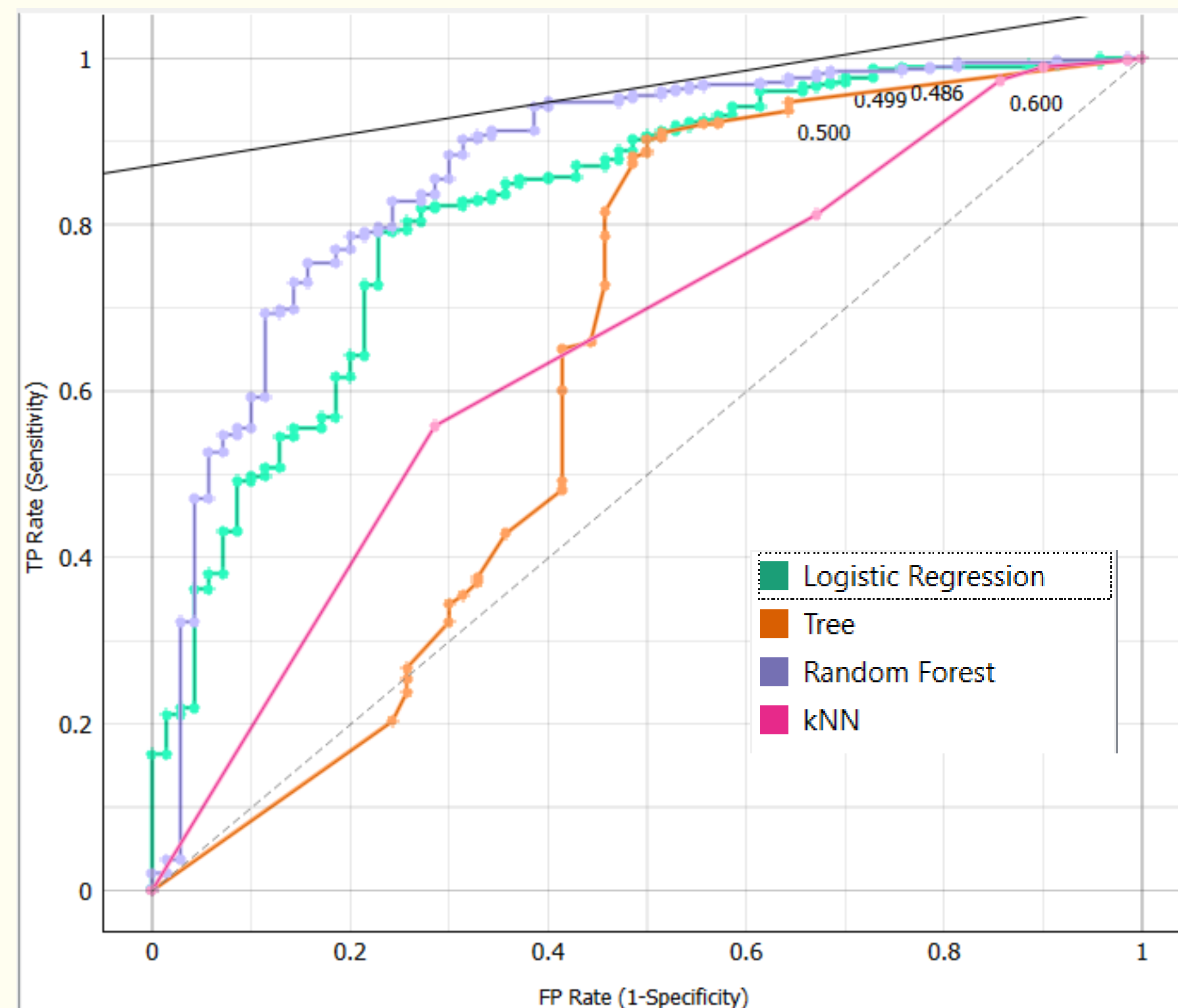
Ejes:

- **X (FP Rate)** → Tasa de falsos positivos.
- **Y (TP Rate)** → Tasa de verdaderos positivos.

Cuanto más cerca está la curva del extremo superior izquierdo, mejor es el rendimiento del modelo.

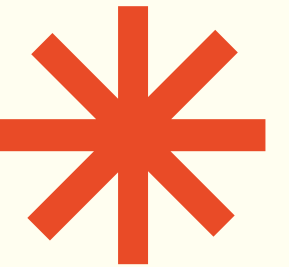
Observaciones:

- **Random Forest (curva violeta)** muestra el mejor desempeño, con una alta sensibilidad y precisión.
- **Logistic Regression** también ofrece un rendimiento sólido, con buena capacidad de discriminación.
- **Tree y kNN** presentan resultados más débiles, con curvas más cercanas a la diagonal





Conclusiones del modelo (impacto en el negocio)



- El modelo predice con alta precisión a los clientes que no aceptan las campañas, lo que permite reducir costos evitando invertir en segmentos con baja probabilidad de respuesta.
- Sin embargo, algunos clientes que sí aceptarían fueron clasificados como que no lo harían, lo que indica una oportunidad de mejora para aumentar la detección de clientes potenciales.
- En general, el modelo identifica patrones útiles para segmentar clientes y orientar las campañas hacia quienes muestran mayor predisposición a aceptar.
- Su aplicación permitiría optimizar el presupuesto de marketing, mejorar la efectividad de las campañas y aumentar el retorno sobre la inversión (ROI).



FIN