

Campaña de Marketing

Inés Murtagh

Objetivo

Entrenar un modelo predictivo que permita a la empresa maximizar el beneficio de la próxima campaña de marketing. Poder predecir quién responderá a una oferta de un producto.





Datos de iFood

consiste en información sobre los clientes de la empresa de delivery de comida, y las interacciones de cada uno de ellos con las campañas enviadas.



29

variables

de las cuales 26 son numéricas y 3 son del tipo carácter.

2240

registros

`AcceptedCmp1` - 1 si el cliente aceptó la oferta en la 1ª campaña, 0 en caso contrario

...

`AcceptedCmp5` - 1 si el cliente aceptó la oferta en la 5ª campaña, 0 en caso contrario

`MntFishProducts` - cantidad gastada en productos pesqueros en los últimos 2 años

`MntMeatProducts` - cantidad gastada en carne en los últimos 2 años

`MntFruits` - cantidad gastada en frutas en los últimos 2 años

`MntSweetProducts` - cantidad gastada en dulces en los últimos 2 años

`MntWines` - cantidad gastada en vino en los últimos 2 años

`NumDealsPurchases` - número de compras realizadas con descuento

`NumCatalogPurchases` - número de compras realizadas con el catálogo

`NumStorePurchases` - número de compras realizadas en el lugar

`NumWebPurchases` - número de compras realizadas con la página web

`NumWebVisitsMonth` - número de visitas al sitio web de la empresa en el último mes

`Complain` - 1 si el cliente se quejó en los últimos 2 años, 0 en caso contrario

`Recency` - número de días desde la última compra

La variable `Response` es la variable target.

Vale 1 si el cliente aceptó la oferta en la última campaña, 0 en caso contrario.

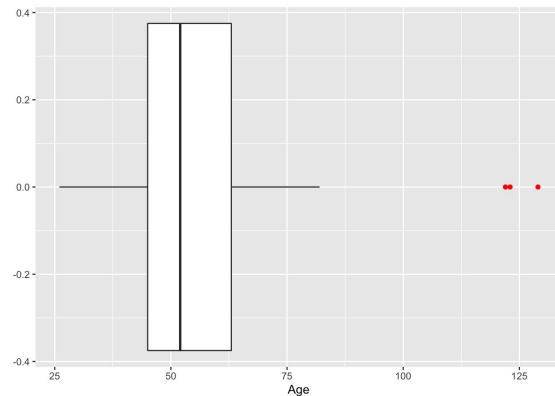


Cambios en el Dataset

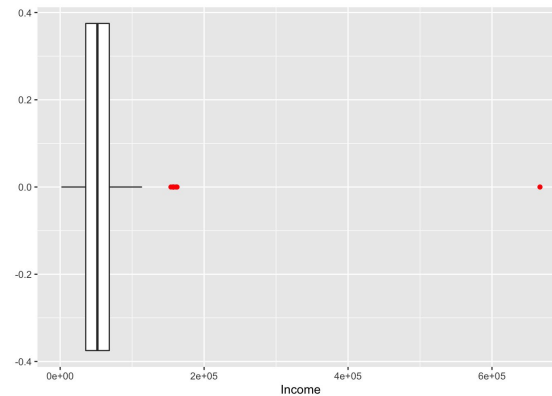
01	Verificar el formato	<ul style="list-style-type: none">• ID como factor, y no como número• Dt_Customer como fecha
02	Creación de nuevas variables	<ul style="list-style-type: none">• Edad (con la fecha de nacimiento)• Hijos: Teens + Kids• Purchases y Products• Accepted
03	Eliminar variables	<ul style="list-style-type: none">• Z_CostContact• Z_Revenue
04	Arreglar las columnas (rename, mutate)	<ul style="list-style-type: none">• Marital Status (4 categorías)• Response
05	Análisis de datos faltantes	<ul style="list-style-type: none">• Income



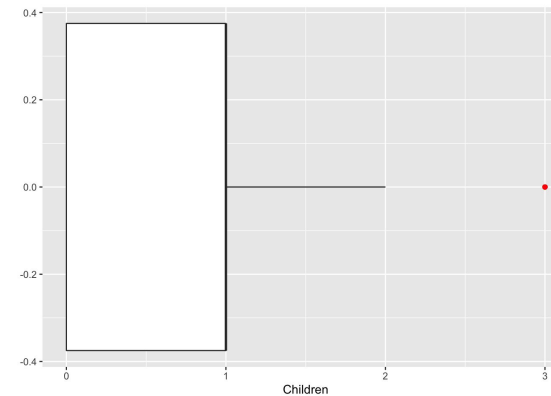
Outliers (valores atípicos)



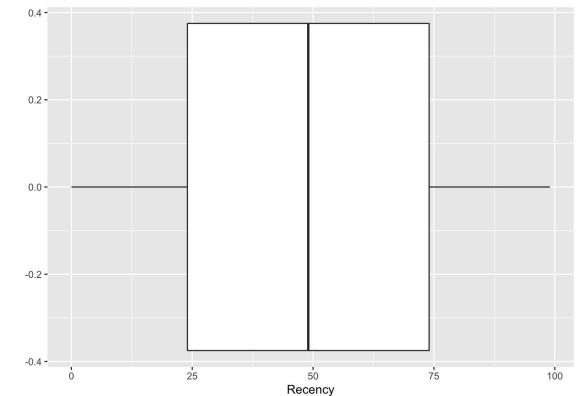
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
26.0	45.0	52.0	53.2	63.0	129.0



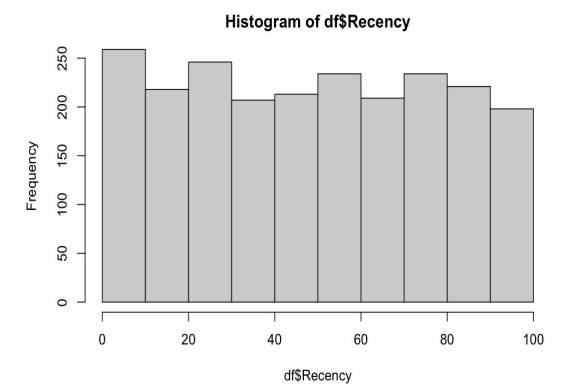
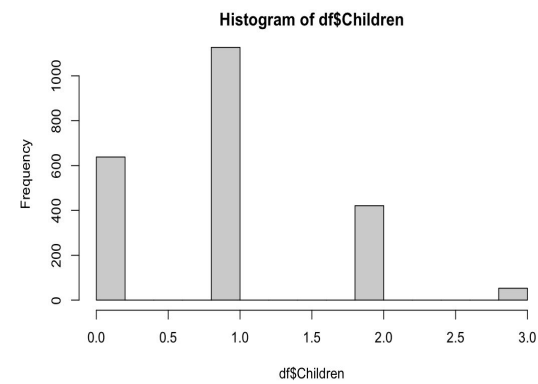
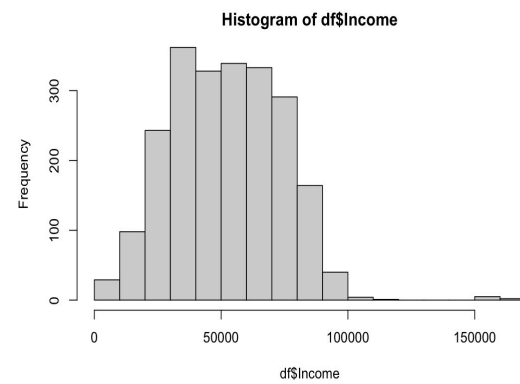
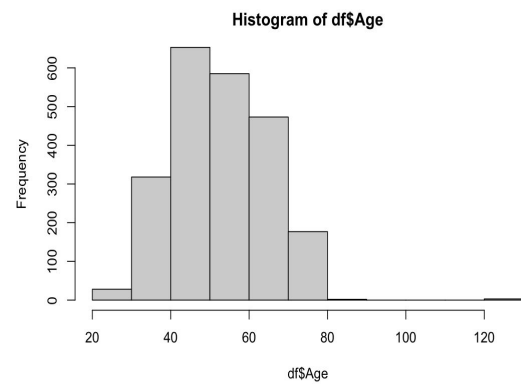
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1730	35539	51742	52247	68290	666666



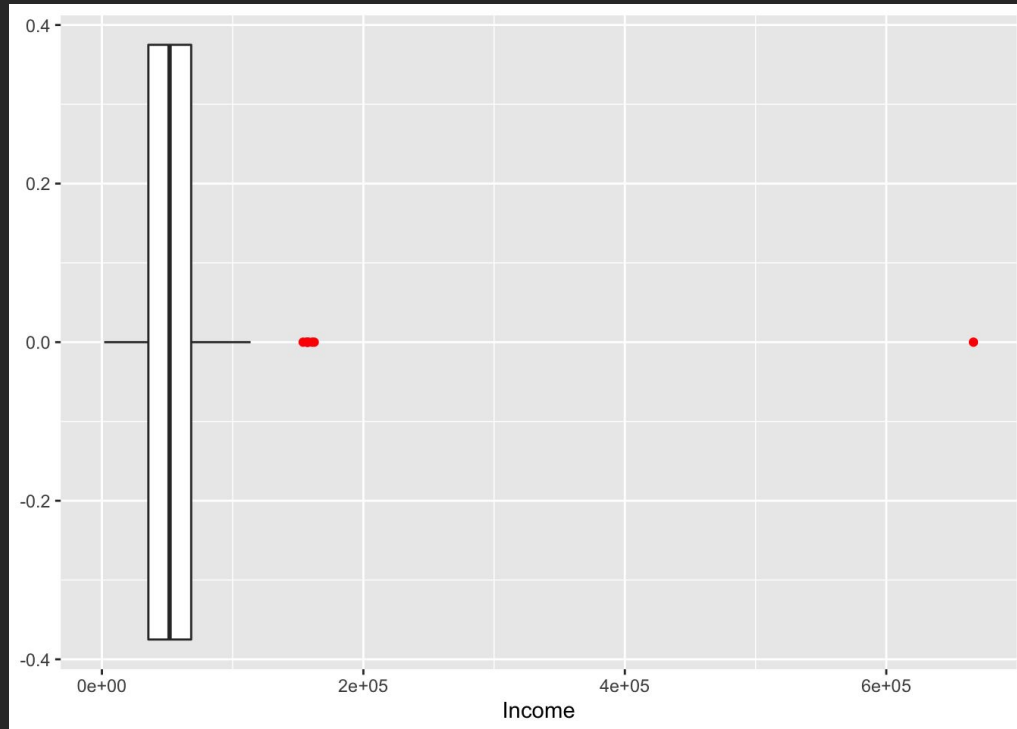
Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.0000	0.0000	1.0000	0.9504	1.0000	3.0000



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
0.00	24.00	49.00	49.12	74.00	99.00



Outliers (valores atípicos)



Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
1730	35539	51742	52247	68290	666666

Education: Graduation

Marital_Status: Graduation

Income: 666666

Enrolment: 2013-06-02

Recency: 23

NumWebVisitsMonth: 6

Complain: 0

Response: No

Age: 45

Children: 1

Purchases: 11

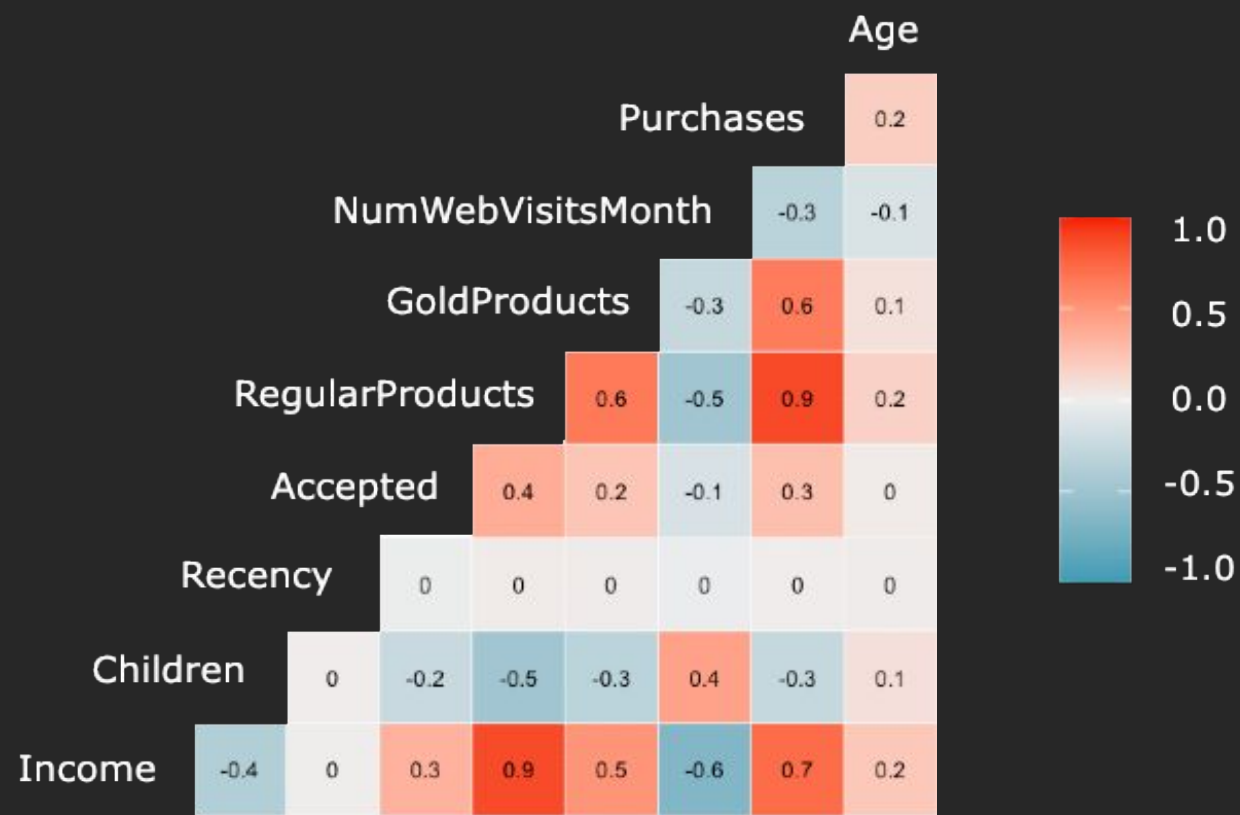
Accepted: 0

RegularProducts: 50

GoldProducts: 12

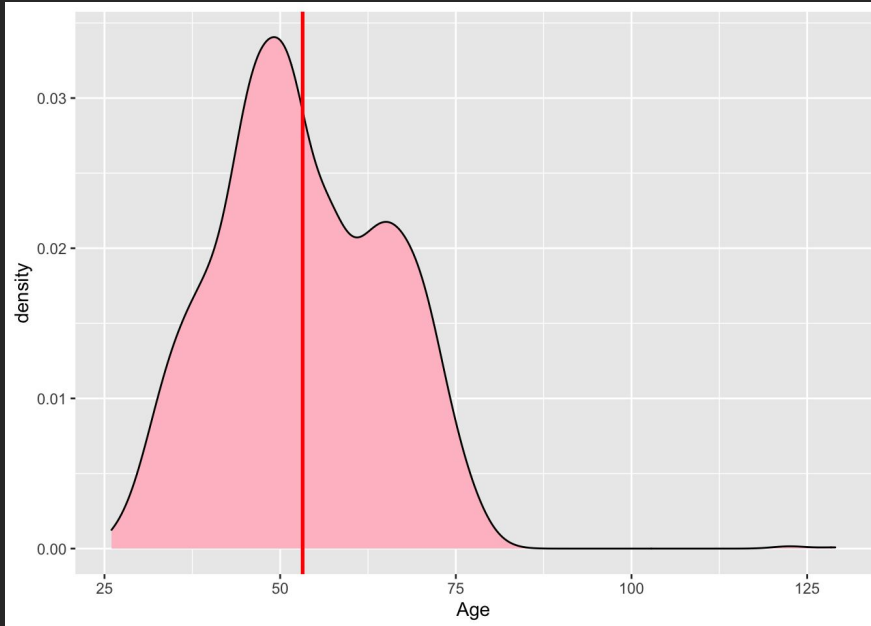
Análisis de correlación

Las variables correlacionadas positivamente están en rojo y las variables negativamente correlacionadas están en azul.



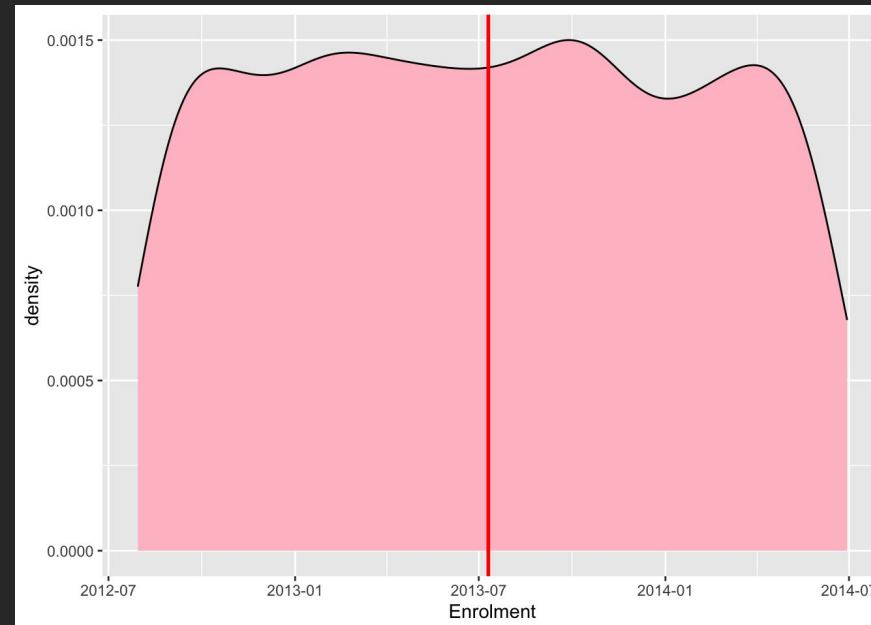
Conocer al cliente

Cantidad por edad



Cómo están distribuidas las edades
Entender cuántos años tienen mis clientes
Entender que público es mi target

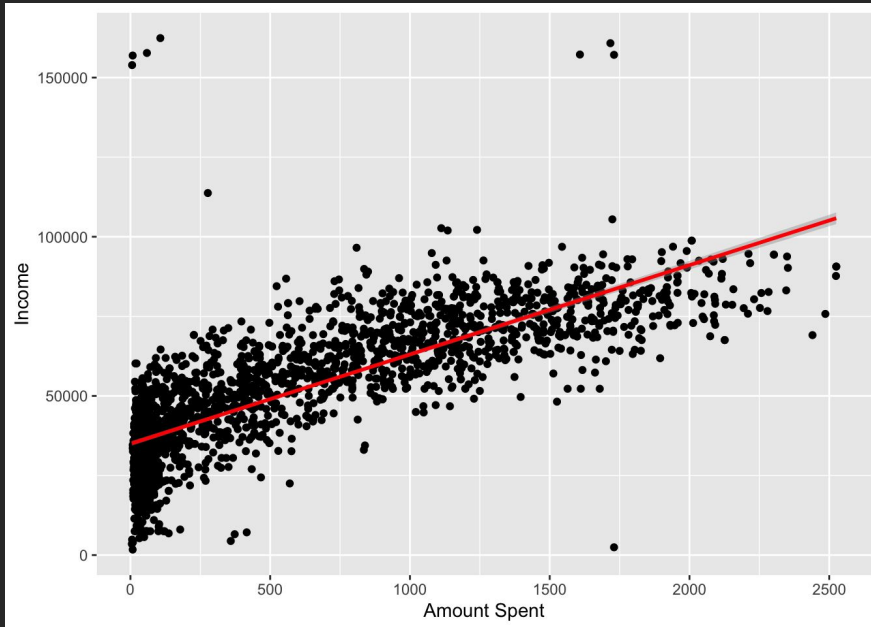
Qué día se dio de alta el cliente
Entender que meses se registran + personas



Fecha de registro

Análisis de la cantidad gastada en productos

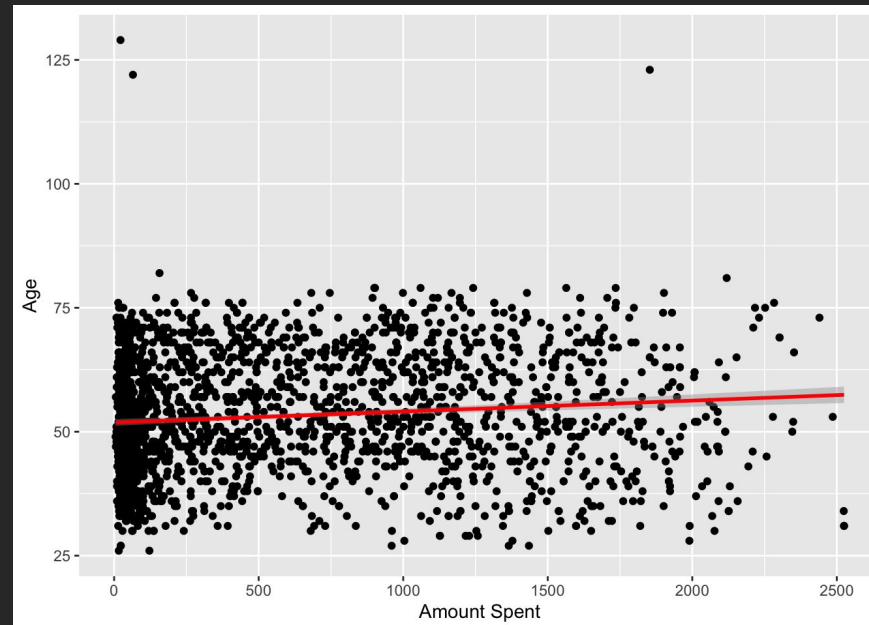
Cantidad gastada por Ingreso



Cómo se relacionan las variables de Ingreso y Cantidad total gastada en productos

¿Existe una relación entre las dos?

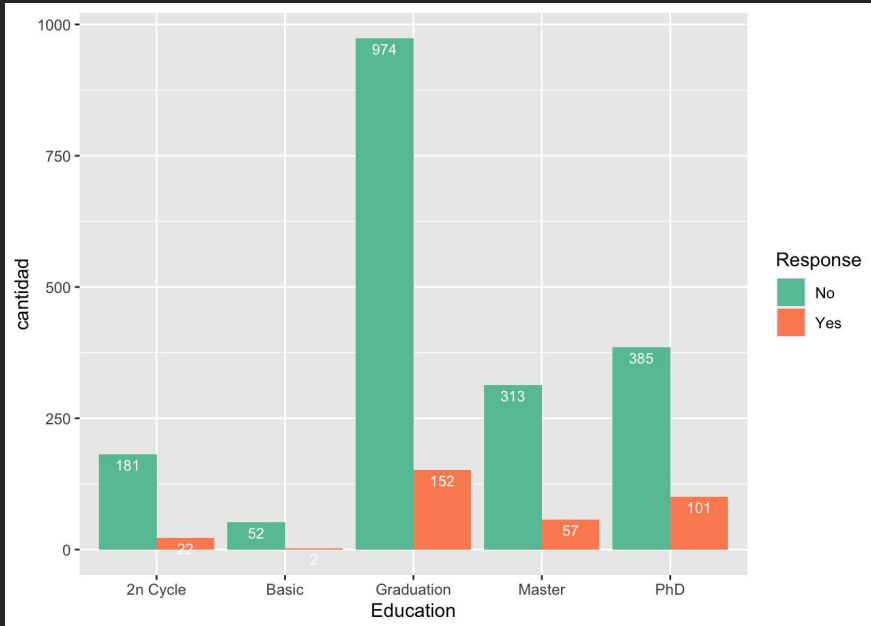
¿Existe una relación entre la edad que tiene una persona, y la cantidad que gasta en productos?



Cantidad gastada por edad

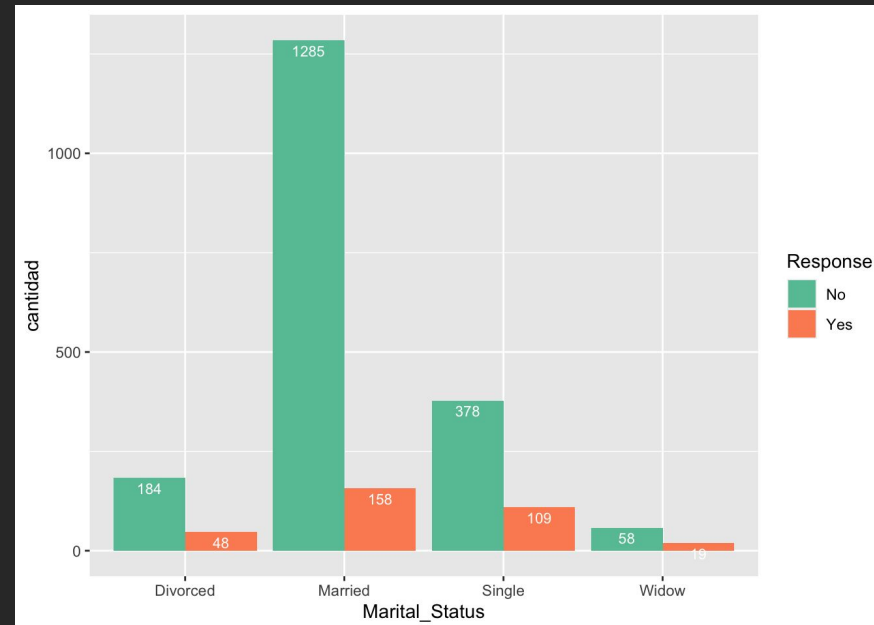
Aceptación de las campañas

Cantidad de respuestas por nivel de Educación



Cuántas personas respondieron, por sí o por no, a una campaña lanzada por la empresa, dependiendo del nivel educativo en el que están

Visualizar por cada tipo de estado civil de una persona, si acepta o no la campaña

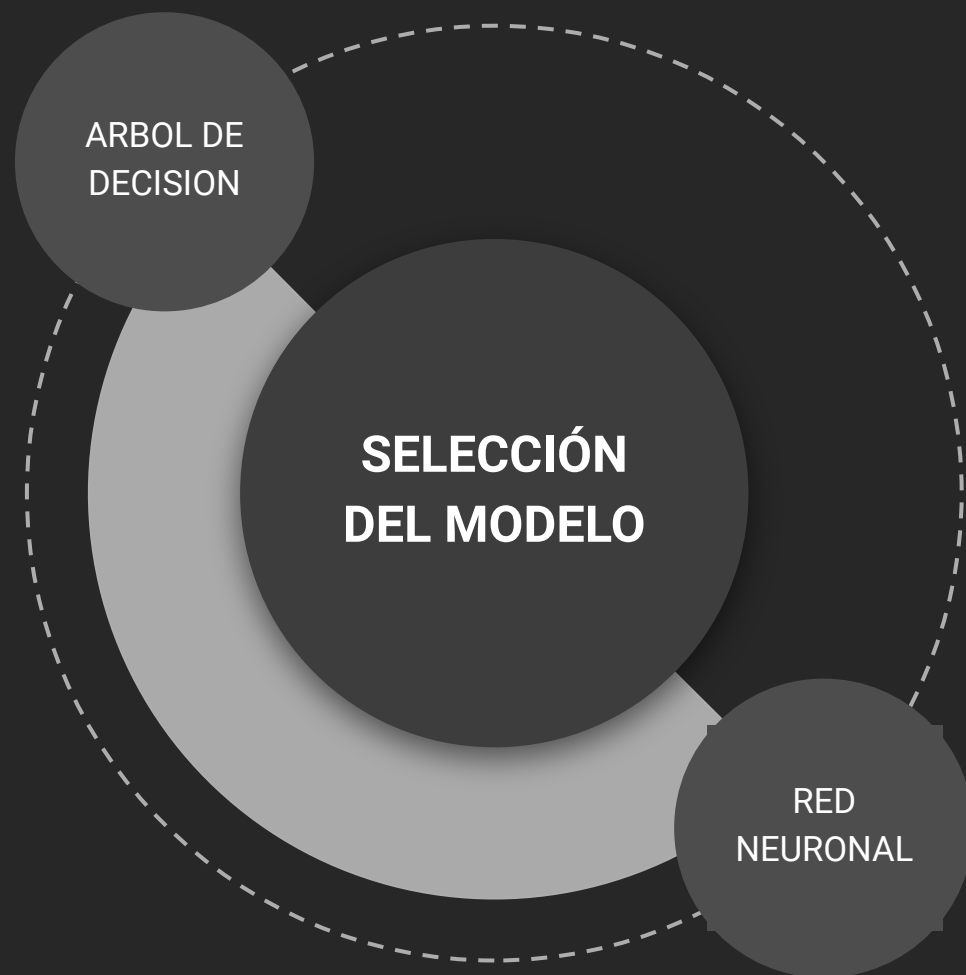


Cantidad de respuestas por estado civil

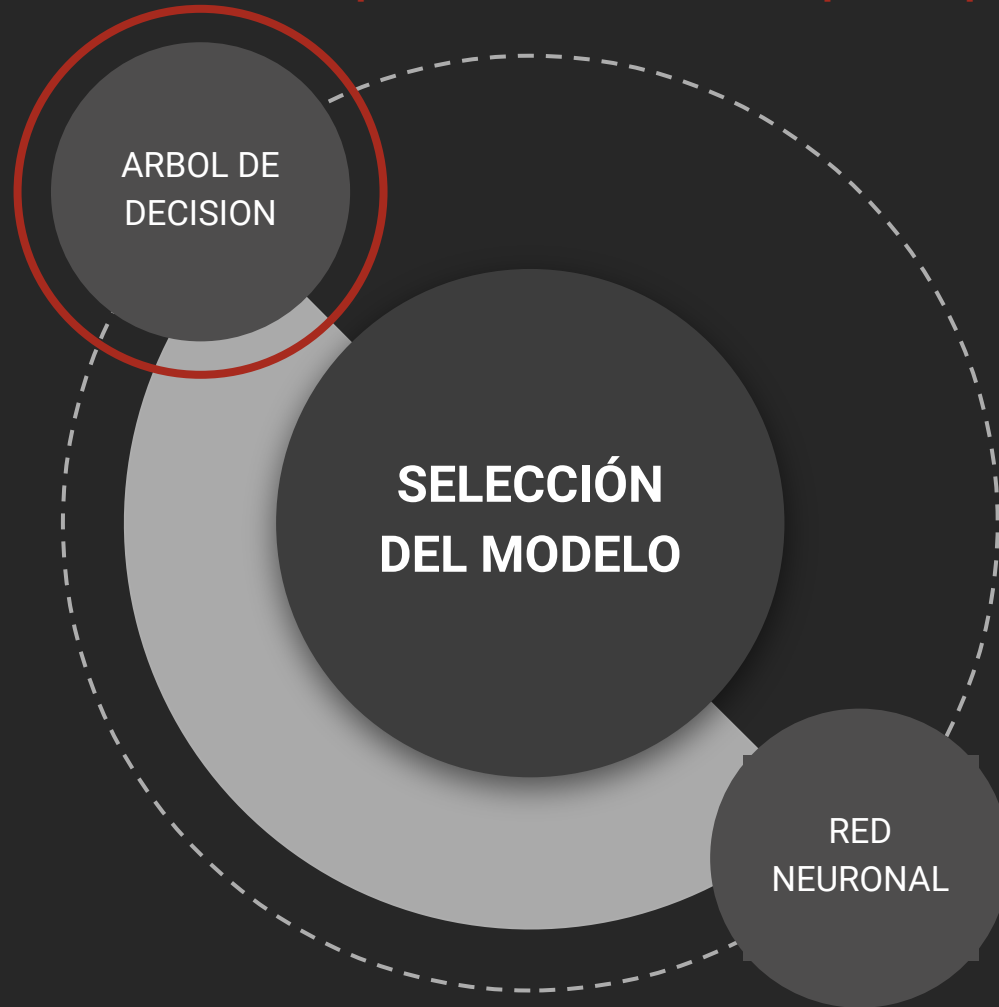
Armado del modelo

predicción





Se elige el árbol de decisión ya que este expresa las variables y condiciones que va a utilizar para la clasificación, lo cual puede ser útil para entender la razón por la que se está aceptando o no la campaña.





	Marital_Status	Income	Recency	NumWebVisitsMonth	Complain	Response
3	Married	71613	26	4	0	No
5	Married	58293	94	5	0	No
6	Married	62513	16	6	0	No
7	Divorced	55635	34	6	0	No
8	Married	33454	32	8	0	No
9	Married	30351	19	9	0	Yes
10	Married	5648	68	20	0	No
11	Married	52247	11	7	0	No
14	Divorced	59354	53	6	0	No
17	Married	41850	51	8	0	No
18	Married	37760	20	7	0	No
19	Married	76995	91	5	0	No
20	Single	33812	86	6	0	No
21	Married	37040	41	8	0	No
22	Married	2447	42	1	0	No

Showing 1 to 14 of 1,792 entries, 11 total columns

	Marital_Status	Income	Recency	NumWebVisitsMonth	Complain	Response
1	Single	58138	58	7	0	Yes
2	Single	46344	38	5	0	No
4	Married	26646	26	6	0	No
12	Married	7500	59	8	0	No
13	Divorced	63033	82	2	0	No
15	Married	17323	38	8	0	No
16	Single	82800	23	3	0	Yes
31	Divorced	10979	34	5	0	No
34	Divorced	46610	8	6	0	Yes
39	Widow	42429	99	5	0	No
41	Married	80011	3	4	0	No
42	Married	20559	88	8	0	No
50	Married	72550	39	8	0	No
64	Widow	67680	8	5	0	No
65	Married	70666	29	4	0	No

Showing 1 to 14 of 447 entries, 11 total columns

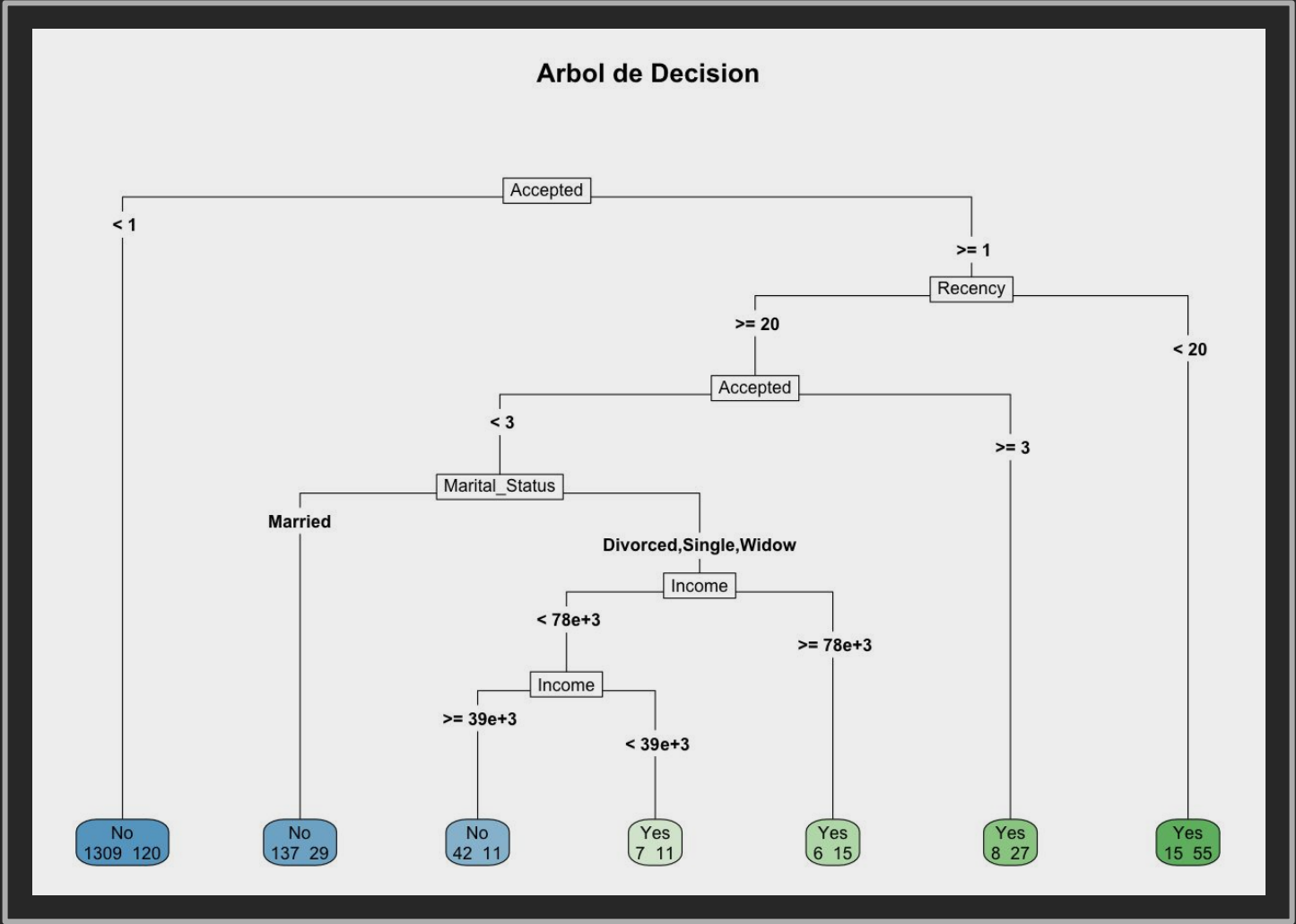
```
set.seed(154); particion=createDataPartition(y=dfpred$Response, p=0.80, list=F)
entreno = dfpred[particion,]
testeo = dfpred[-particion,]
```

Quedaron 1792 registros en entreno, y 447 en testeo

En el árbol de decisión de tipo clasificación se utilizan todas las variables menos ID, Fecha, Educación y Edad

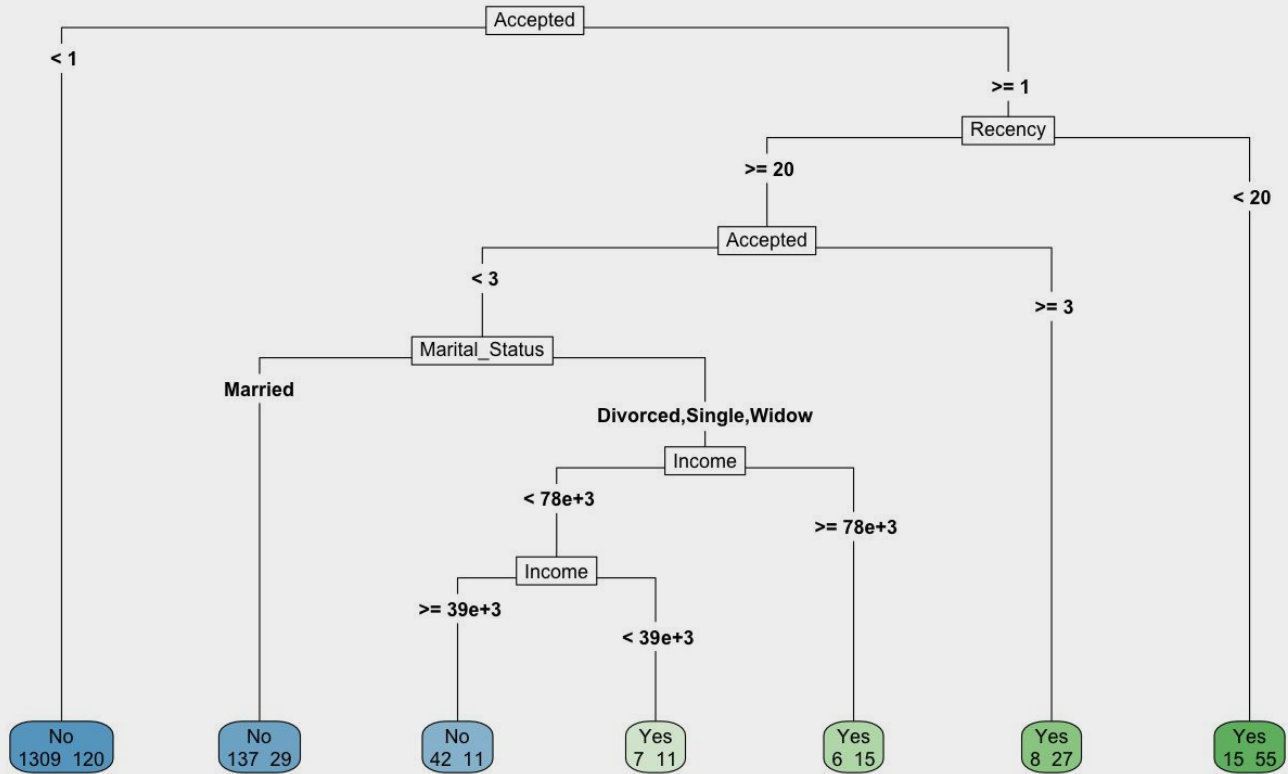
		Reference	
Prediction		No	Yes
	No	366	39
	Yes	15	27

Accuracy: 0.8792 (87,9%)



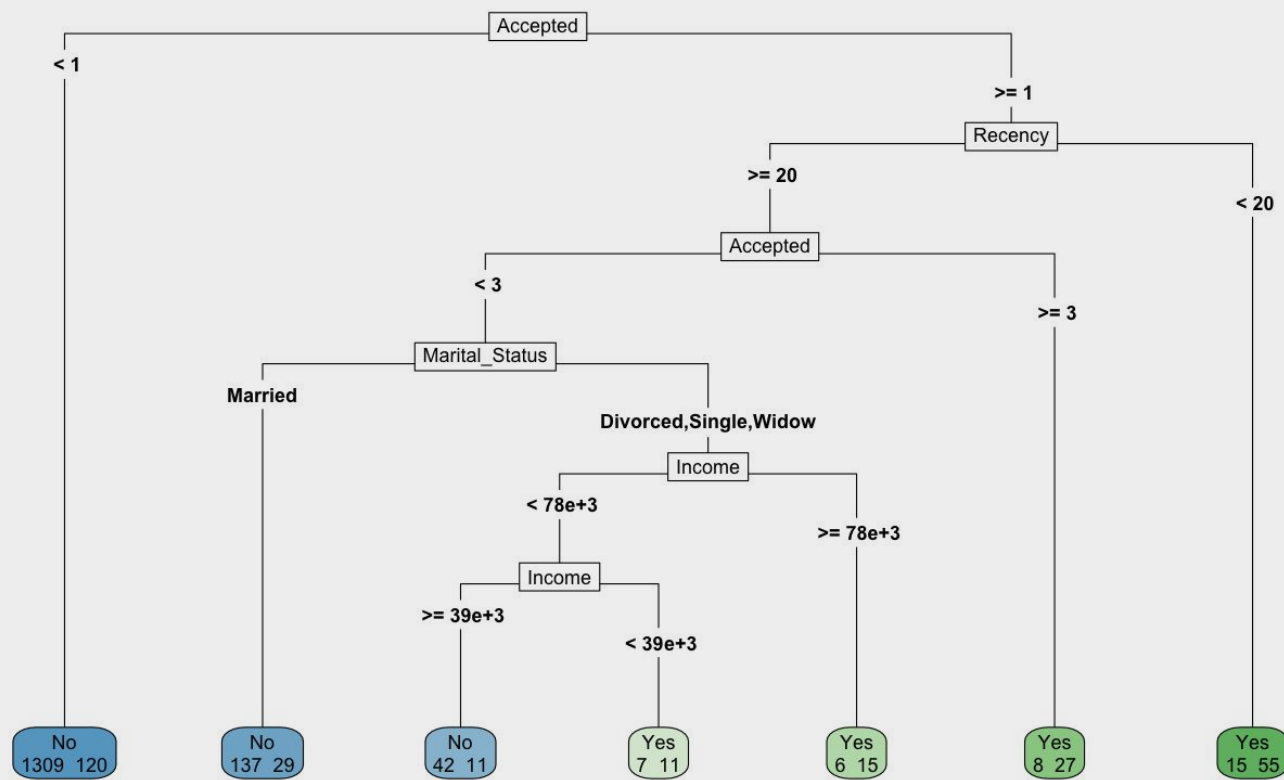
	ID	Education	Marital_Status	Income	Enrolment	Recency	NumWebVisitsMonth	Complain	Response	Age	Children	Purchases	Accepted
418	5067	Graduation	Married	80134	2014-02-14	11	5	0	No	28	0	21	4

Arbol de Decision



	ID	Education	Marital_Status	Income	Enrolment	Recency	NumWebVisitsMonth	Complain	Response	Age	Children	Purchases	Accepted
418	5067	Graduation	Married	80134	2014-02-14	11	5	0	No	28	0	21	4

Arbol de Decision



```

> fila = df[418,]
> predict(arbol, fila, type="class")
418
Yes
Levels: No Yes

```

El equipo de marketing necesita mejorar su estrategia utilizada para adquirir nuevos clientes, y una buena manera de empezar es a través de una base de datos como esta.

Analizar el conjunto de datos es clave para comprender los problemas y proponer soluciones basadas en datos.

- Medir la participación de los clientes en las campañas de marketing
- Evaluar el éxito de campañas anteriores
- Proponer soluciones basadas en datos

El desafío para los equipos de marketing es saber a quién están tratando de llegar, y esto se hace contando con la información sobre cada uno de los clientes.

Hay un gran valor en saber cómo apelar a los distintos usuarios que una organización tiene. Que cada una pueda recibir un mensaje diferente que se relaciona más con ellos, viniendo de la misma campaña de marketing.

Conclusiones

