

Realizado por:

Juan Camilo Restrepo Velez  
Wilder Valencia Ocampo

## **PRÁCTICA DE PREPARACIÓN DE DATOS 10%**

Seleccionar una base de datos en <https://www.datos.gov.co/> y realizar un informe con todos los pasos de preparación de datos, utilizar pantallazos para documentar los resultados. Después de realizar todos los pasos, responder en el informe:

### **Bank Marketing**

Los datos están relacionados con campañas de marketing directo de una institución bancaria portuguesa. Las campañas de marketing se basaron en llamadas telefónicas. A menudo, se requería más de un contacto con el mismo cliente, para poder acceder a si el producto (depósito bancario a plazo) sería ('sí') o no ('no') suscrito.

### **Información de atributos**

Información Bancaria de los clientes

Age - Edad

Job - Trabajo: tipo de trabajo

Marital - Estado civil: estado civil

Education - Educación: Nivel educativo

Default - Incumplimiento: ¿tiene el crédito en mora?

Housing - Vivienda: ¿tiene un préstamo de vivienda?

Loan - Préstamo: ¿tiene préstamo personal?

Relacionado con la última llamada de la actual campaña

Contact - Contacto: tipo de comunicación

Month - Mes: último mes de contacto del año

DayofWeek - Día de la semana: último día de contacto de la semana

Duration - Duración: duración del último contacto, en segundos (numérico). Nota importante: este atributo afecta en gran medida al objetivo de salida (por ejemplo, si la duración = 0, entonces y = "no"). Sin embargo, no se conoce la duración antes de una llamada se realiza. Además, después del final de la llamada se conoce obviamente y. Por lo tanto, esta entrada sólo debe incluirse a efectos de referencia y debe descartarse si se pretende tener un modelo predictivo realista.

## Otros

Campaign - Campaña: número de contactos realizados durante esta campaña y para este cliente

Pdays - pDías: número de días que pasaron después de que el cliente fue contactado por última vez en una campaña anterior. Nota, 999 significa que el cliente no fue contactado anteriormente

Previous - Anterior: número de contactos realizados antes de esta campaña y para este cliente

Poutcome: resultado de la anterior campaña de marketing

## Atributos del contexto social y económico

Emp.var.rate - Tasa de variación del empleo - indicador trimestral

Cons.price.idx: Índice de Precios al Consumidor - Indicador mensual; el Índice de Precios al Consumidor o IPC mide los cambios en los precios pagados por los consumidores por una cesta de bienes y servicios cada mes.

Cons.conf.idx: Índice de confianza del consumidor - Indicador mensual; En Portugal, el índice de confianza del consumidor se basa en entrevistas con los consumidores sobre sus percepciones de la situación económica actual y futura del país y sus tendencias de compra. Se estima utilizando la diferencia entre la proporción de respuestas de evaluación positivas y las respuestas de evaluación negativas, pero no incluye la proporción de respuestas neutras

Euribor3m: euribor 3 meses - Euribor es la abreviatura de Euro Interbank Offered Rate. es un índice de referencia publicado diariamente que indica el tipo de interés promedio al que un gran número de bancos europeos dicen concederse préstamos a corto plazo entre ellos para prestárselo a terceros.

Nr.employed - Número de empleados: Número de empleados - Indicador trimestral; Número de personas empleadas para el trimestre.

## Variable objetivo

y - ¿el cliente ha suscrito un depósito a plazo?

\*Tomado de <https://www.kaggle.com/henriqueyamahata/bank-marketing>

La variable **pdays** resulta irrelevante desde reglas del negocio,

**Selected attribute**

Name: pdays  
Missing: 0 (0%)  
Distinct: 27  
Type: Numeric  
Unique: 4 (0%)

Statistic	Value
Minimum	0
Maximum	999
Mean	962.475
StdDev	186.911

Class: y (Nom)

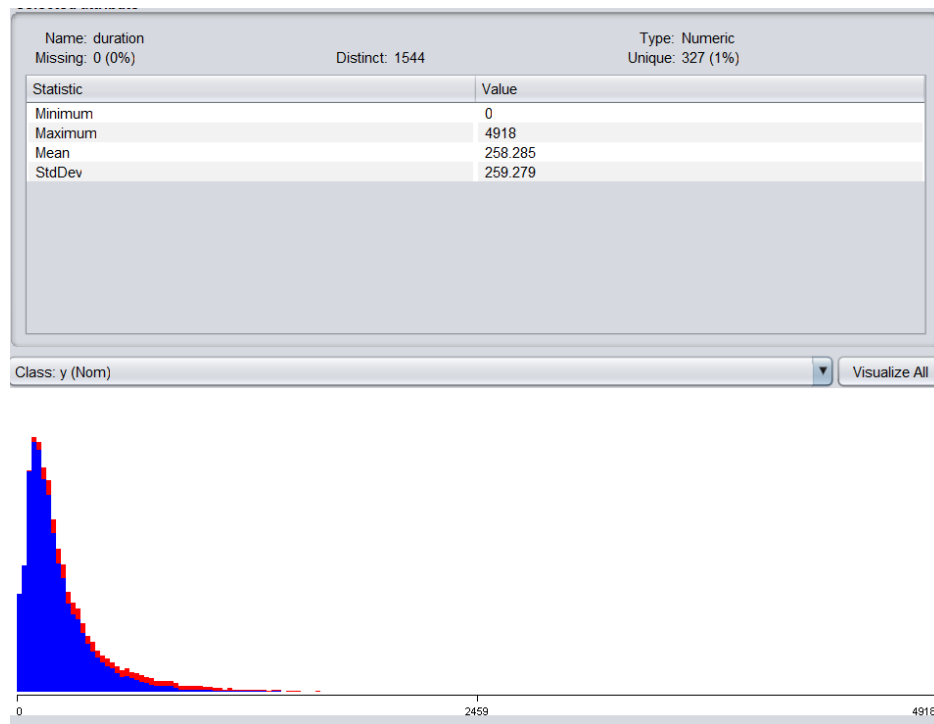
Name: contact  
 Missing: 0 (0%)  
 Distinct: 2  
 Type: Nominal  
 Unique: 0 (0%)

No.	Label	Count	Weight
1	telephone	15044	15044.0
2	cellular	26144	26144.0

Class: y (Nom)

## 2. Cuáles variables son redundantes?

La variable **duration** resulta redundante desde reglas del negocio, porque si la duración es igual 0, entonces la variable objetivo es "no".



Paso extra -> Estadística descriptiva (Paso 3)



### **3. Cuáles son los datos atípicos?**

No hay presencia de atípicos. Esto se debe en gran medida a que los datos ya tenían una alta calidad, por su procedencia de una institución como lo es un banco.

#### 4. Cuáles son los datos nulos?

Varias variables contaban con la categoría “unknown”, por lo cual desde la sabana de datos se procedió a pasar esta categoría a nula.

Variables con Missing							
Name: job Missing: 330 (1%)      Distinct: 11      Type: Nominal Unique: 0 (0%)				Name: education Missing: 1731 (4%)      Distinct: 7      Type: Nominal Unique: 0 (0%)			
No.	Label	Count	Weight	No.	Label	Count	Weight
1	housemaid	1060	1060.0	1	basic.4y	4176	4176.0
2	services	3969	3969.0	2	high.school	9515	9515.0
3	admin.	10422	10422.0	3	basic.6y	2292	2292.0
4	blue-collar	9254	9254.0	4	basic.9y	6045	6045.0
5	technician	6743	6743.0	5	professional.c...	5243	5243.0
6	retired	1720	1720.0	6	university.degr...	12168	12168.0
7	management	2924	2924.0	7	illiterate	18	18.0
8	unemployed	1014	1014.0				
9	self-employed	1421	1421.0				
10	entrepreneur	1456	1456.0				
11	student	875	875.0				
Name: marital Missing: 80 (0%)      Distinct: 3      Type: Nominal Unique: 0 (0%)				Name: default Missing: 8597 (21%)      Distinct: 2      Type: Nominal Unique: 0 (0%)			
No.	Label	Count	Weight	No.	Label	Count	Weight
1	married	24928	24928.0	1	no	32588	32588.0
2	single	11568	11568.0	2	yes	3	3.0
3	divorced	4612	4612.0				
Name: loan Missing: 990 (2%)      Distinct: 2      Type: Nominal Unique: 0 (0%)							
No.	Label	Count	Weight				
1	no	33950	33950.0				
2	yes	6248	6248.0				

## 5. Cuál es el valor de la imputación?

Name: job Missing: 0 (0%) Distinct: 11 Type: Nominal Unique: 0 (0%)				Name: education Missing: 0 (0%) Distinct: 7 Type: Nominal Unique: 0 (0%)			
No.	Label	Count	Weight	No.	Label	Count	Weight
1	housemaid	1060	1060.0	1	basic.4y	4176	4176.0
2	services	3969	3969.0	2	high.school	9515	9515.0
3	admin.	10752	10752.0	3	basic.6y	2292	2292.0
4	blue-collar	9254	9254.0	4	basic.9y	6045	6045.0
5	technician	6743	6743.0	5	professional.c...	5243	5243.0
6	retired	1720	1720.0	6	university.degr...	13899	13899.0
7	management	2924	2924.0	7	illiterate	18	18.0
8	unemployed	1014	1014.0				
9	self-employed	1421	1421.0				
10	entrepreneur	1456	1456.0				
11	student	875	875.0				

admin

Name: marital Missing: 0 (0%) Distinct: 3 Type: Nominal Unique: 0 (0%)				Name: default Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)			
No.	Label	Count	Weight	No.	Label	Count	Weight
1	married	25008	25008.0	1	no	41185	41185.0
2	single	11568	11568.0	2	yes	3	3.0
3	divorced	4612	4612.0				

married

Name: loan Missing: 0 (0%) Distinct: 2 Type: Nominal Unique: 0 (0%)							
No.	Label	Count	Weight				
1	no	34940	34940.0				
2	yes	6248	6248.0				

no

## Paso Extra → Creación nuevas variables (Paso 5)

Aunque no se hizo necesario tener que crear las variables manualmente debido a que el dataset ya contaba con una separación de datos de una variable **date** por dos nuevas variables **month** y **day\_of\_week**:

Name: month		Type: Nominal	
Missing: 0 (0%)		Distinct: 10	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	may	13769	13769.0
2	jun	5318	5318.0
3	jul	7174	7174.0
4	aug	6178	6178.0
5	oct	718	718.0
6	nov	4101	4101.0
7	dec	182	182.0
8	mar	546	546.0
9	apr	2632	2632.0
10	sep	570	570.0

Name: day_of_week		Type: Nominal	
Missing: 0 (0%)		Distinct: 5	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	mon	8514	8514.0
2	tue	8090	8090.0
3	wed	8134	8134.0
4	thu	8623	8623.0
5	fri	7827	7827.0

## 6. Cuáles variables tienen una alta correlación?

**marital=married** con **marital=single** de -0.78 → Al ser variables dummies de una misma variable que tiene más categorías no se puede eliminar ninguna de las dos, además de que no se puede evidenciar una colinealidad entre ellas.

**poutcome=nonexistent** con **previous** de -0.88 → Al existir una alta correlación entre estas variables es posible una redundancia, por lo que se debe eliminar una de las dos variables. En este sentido se analiza la correlación de la variable **previous** con las demás categorías de **poutcome**, en este modo, las correlaciones son de: 0.68 con failure y 0.52 con success, llegando a la conclusión de que se debe eliminar una de las dos variables, **lo cual se decidirá según la que tenga menor correlación con la variable objetivo, que es poutcome con 0.18091**

**poutcome=failure** con **poutcome=nonexistent** de -0.85 → Al ser variables dummies de una misma variable que tiene más categorías no se puede eliminar ninguna de las dos, además de que no se puede evidenciar una colinealidad entre ellas.

**cons.price.idx** con **emp.var.rate** de 0.78 → Al ser variables económicas y de contexto social representan grupos de personas, por lo tanto, presentan redundancia entre ellas. Al aumentar la tasa de empleo, los precios de la canasta básica pueden tender a aumentar debido a la oferta y demanda.

**euribor3m** con **emp.var.rate** de 0.97 → Al ser variables económicas y de contexto social representan grupos de personas, por lo tanto, presentan redundancia entre ellas. Por razones de especulación y de movimientos de libre mercado los intereses que manejan los bancos tienen relación directa en las empresas y en como estas pueden o no contratar.

**nr.employed** con **emp.var.rate** de 0.91 → Al ser variables económicas y de contexto social representan grupos de personas, por lo tanto, presentan redundancia entre ellas. La cantidad de personas contratadas por trimestre influye directamente en la tasa de empleo.

Además, como la misma variable **emp.var.rate** presenta tres correlaciones altas con otras esta es la que se decide eliminar. **También se tiene en cuenta su correlación con la variable objetivo que es de 0.29833**

**nr.employed** con **euribor3m** de 0.95 → Al ser variables económicas y de contexto social representan grupos de personas, por lo tanto, presentan redundancia entre ellas. En este sentido, se relacionan entre la capacidad empleabilidad (**nr.employed**) y la tasa de interés que reciben de otros bancos (**euribor3m**), entre mayor empleabilidad mayor tasa de interés. **Se decidirá cuál eliminar según la correlación que tengan con la variable objetivo, por lo que se decide eliminar esta última variable al presentar una correlación de 0.30777**



	age	job=housemaid	job=services	job=admin	job=blue-collar	job=technician	job=retired	job=management	job=unemployed	job=self-employed	job=entrepreneur	job=student	marital=married	marital=single	marital=divorced	education=basic,4y	education=high,school	education=basic,6y	education=basic,9y	education=professional,course	education=university,degree	education=illiterate	default=yes	housing=yes	loan=yes	month=may	month=jun	month=jul	month=aug	month=oct	month=nov	month=dec	month=mar	month=apr	month=sep	day_of_week=mon	day_of_week=tue	day_of_week=wed	day_of_week=thu	day_of_week=fri	campaign	previous	poutcome=nonexistent	poutcome=failure	poutcome=success	emp,var,rate	cons,price,idx	cons,conf,idx	euribor3m	nr,employed	
age	0	0.09	-0.07	-0.09	-0.02	-0.06	0.44	0.06	0	0	0.03	-0.2	0.27	-0.41	0.17	0.24	-0.11	0.01	-0.04	0	-0.04	0.02	0	0	-0.01	-0.07	-0.01	-0.04	0.07	0.05	0.03	0.05	0.01	0.01	0.04	0.02	0.02	0.02	-0.02	-0.02	0.01	0	0.02	0	0.04	0	0	0.13	0.01	-0.02	
job=housemaid	0.09	0	-0.05	-0.1	-0.09	-0.07	-0.03	-0.04	-0.03	0	0.03	-0.02	0.04	-0.06	0.02	0.19	-0.03	0.01	-0.03	0	-0.03	0	0	0	0	-0.02	0	0.02	0.02	0	-0.01	0.01	-0.01	0	0	0.01	0	0.01	0	-0.01	0	-0.01	0	0.01	0	0.04	0.03	0.04	0.04	0.03	
job=services	-0.07	-0.05	0	-0.19	-0.18	-0.14	-0.07	-0.09	-0.05	-0.06	0	0.02	-0.07	-0.05	0.02	-0.07	-0.18	0	-0.01	-0.03	0.34	0	0	0	0	0.06	0.01	0.02	-0.07	-0.02	-0.02	-0.02	0.01	0	0	0	0	0	0	0	0	0	0	0.03	0.02	0.01	0.01	0.02			
job=admin	-0.09	-0.1	-0.19	1	-0.32	-0.26	-0.12	-0.16	-0.09	-0.11	-0.11	-0.09	-0.12	-0.11	-0.16	-0.18	0.12	-0.01	-0.16	0	-0.16	-0.02	-0.01	0.01	0.02	-0.05	-0.01	0	0.08	0.01	0	0	0.01	-0.01	0.01	0	0	0	0.01	0.01	0.02	-0.01	0	0.03	-0.02	-0.04	0.04	-0.02	-0.02		
job=blue-collar	-0.02	-0.09	-0.18	-0.32	1	-0.24	-0.11	-0.15	-0.09	-0.1	-0.1	-0.08	0.13	-0.1	-0.06	0.27	-0.17	0.23	0.37	-0.13	-0.32	0.01	0	-0.01	-0.01	0.14	-0.03	0.03	-0.13	-0.05	-0.06	-0.03	-0.04	-0.01	-0.05	-0.01	-0.01	0.02	-0.01	0	-0.05	0.04	-0.01	-0.06	0.06	0.08	-0.1	0.05	0.06		
job=technician	-0.06	-0.07	-0.14	-0.26	-0.24	1	-0.09	-0.12	-0.07	-0.08	-0.08	-0.07	-0.06	0.06	0	-0.14	-0.11	-0.11	0.48	-0.04	-0.01	0.01	0.01	-0.01	-0.06	-0.03	-0.02	0.14	0	-0.01	-0.01	-0.02	0	0	0	-0.01	0.01	-0.02	0	-0.01	0	-0.02	-0.01	0.05	-0.01	0.05	0.05	0.05			
job=retired	0.44	-0.03	-0.07	-0.12	-0.11	-0.09	1	-0.06	-0.03	-0.04	-0.04	-0.03	0.06	-0.11	0.06	0.17	-0.03	-0.01	-0.04	0.01	-0.03	-0.01	0	-0.01	-0.06	-0.02	-0.01	0.03	0.09	-0.01	0.05	0.04	0.02	0.06	0	0.01	0	-0.01	0	-0.01	0.07	-0.05	0.02	0.07	-0.1	-0.05	0.09	-0.1	-0.13		
job=management	0.06	-0.04	-0.09	-0.16	-0.15	-0.12	-0.06	1	-0.04	-0.05	-0.05	-0.04	0.06	-0.07	0	-0.06	-0.08	-0.03	-0.07	-0.08	-0.02	-0.01	0	-0.01	0	-0.01	-0.01	-0.03	-0.02	0	0.09	0	0	0	0.01	0	-0.01	-0.01	0.01	0.01	0	-0.02	-0.03	0	0	0	0				
job=unemployed	0	-0.03	-0.05	-0.09	-0.09	-0.07	-0.03	-0.04	1	-0.03	-0.03	-0.02	0.01	-0.01	0.01	0	0.01	-0.02	0.02	0.01	-0.02	0.02	0.02	0.01	0	-0.02	0.02	0	-0.01	0.01	0.03	0.01	0.01	-0.01	0.02	-0.01	0	-0.01	0	-0.01	0.01	-0.01	0.02	-0.02	0	0.02	-0.01	-0.02			
job=self-employed	0	-0.03	-0.06	-0.11	-0.1	-0.08	-0.04	-0.05	-0.03	1	-0.04	-0.03	0.01	-0.01	-0.01	-0.02	-0.03	0	-0.01	-0.07	-0.03	0	-0.01	0	-0.01	-0.02	0.01	-0.01	0	0	0.03	0	0.01	0	0	0.01	-0.01	0	-0.01	0	-0.01	0	-0.01	0	0.01	0.01	0.01				
job=entrepreneur	0.03	-0.03	-0.06	-0.11	-0.1	-0.08	-0.04	-0.05	-0.03	-0.04	0.01	0	-0.03	0.01	0	-0.02	-0.01	0.05	-0.01	0.05	0	0.01	0	0	-0.01	-0.02	0.01	-0.05	-0.01	-0.05	-0.02	-0.01	-0.02	0.01	0.01	-0.01	0.01	-0.01	0	-0.01	0	-0.01	0	-0.01	-0.03	0.01	0.01	0.01			
job=student	-0.2	-0.02	-0.05	-0.09	-0.1	-0.08	-0.04	-0.05	-0.03	-0.04	-0.03	-0.03	0.05	-0.06	-0.05	-0.03	-0.03	-0.01	-0.03	-0.05	-0.03	-0.03	-0.03	0	0	-0.02	-0.01	-0.02	-0.01	-0.05	-0.02	0.04	-0.02	0.03	0.05	0	0	0	0	0	0	-0.02	0.14	-0.06	0.01	-0.15	-0.17				
marital=married	0.27	0.04	-0.02	-0.12	0.13	-0.06	0.06	0.06	0.01	0.01	0.05	-0.17	1	-0.78	1	-0.22	0.12	-0.07	0.08	0.07	-0.07	0	0	0.01	0.01	-0.01	0	0.02	-0.02	-0.05	0.04	-0.01	0.01	0	-0.04	-0.02	-0.01	0	0	-0.01	0.01	0	-0.04	0.09	0.09	0.08					
marital=single	-0.41	-0.06	0	0.11	-0.1	0.06	-0.11	-0.07	-0.01	-0.01	-0.06	0.22	-0.78	1	-0.22	-0.13	0.06	-0.07	-0.06	-0.01	-0.11	-0.01	-0.01	0.01	0	-0.02	-0.02	0.04	-0.02	0.02	-0.02	0	0.05	0.01	0.01	-0.01	-0.01	0.01	0.02	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	-0.1				
marital=divorced	0.17	0.02	0.02	0.02	-0.06	0	0.06	0	0.01	-0.01	0.01	-0.05	-0.44	-0.22	1	0.01	0.02	-0.03	-0.02	-0.02	-0.01	0.11	-0.01	-0.01	0	-0.01	-0.01	0.02	-0.03	-0.01	0.02	0	0.01	-0.01	0.01	0	-0.01	0.01	0	-0.01	0.02	0.02	-0.02	0.02	0.02	0.02					
education=basic,4y	0.24	0.19	-0.07	-0.18	0.27	-0.14	0.17	-0.06	0	-0.02	0	-0.03	0.12	-0.13	0.01	1	-0.18	-0.13	-0.13	-0.24	-0.01	0	-0.01	0	-0.01	0	0.02	0.02	0	-0.03	0.01	-0.03	0	0	0	0	0	0	0	-0.01	0.03	0.05	0.02	0.03	0.01						
education=high,school	-0.11	-0.03	0.34	0.11	-0.17	-0.11	-0.03	-0.08	0.01	-0.07	-0.03	0.06	-0.07	0.06	0.02	-0.18	1	-0.13	-0.23	-0.21	-0.39	-0.01	0	-0.01	0	0.04	0	0.03	-0.07	0	-0.01	0	-0.02	0	-0.01	0.01	0	-0.02	0.03	-0.01	-0.02	0.03	-0.01	-0.02	-0.02						
education=basic,6y	0.01	0.01	0	-0.1	0.23	-0.08	-0.01	-0.03	-0.02	-0.03	-0.01	-0.03	0.08	-0.07	-0.03	-0.08	-0.13	1	-0.1	-0.09	-0.17	-0.01	0	-0.01	-0.01	0.06	0.01	0.01	-0.06	-0.01	-0.02	-0.02	-0.02	-0.01	-0.01	-0.01	0	0.01	0	-0.02	0.02	-0.01	-0.02	0.03	-0.03	0.02	0.03				
education=basic,9y	-0.04	-0.03	-0.05	-0.16	0.37	-0.11	-0.04	-0.07	0.02	0	0	-0.01	0.07	-0.06	-0.02	-0.14	-0.23	-0.1	-0.3	-0.16	-0.3	-0.01	0	-0.01	0.09	0.01	0.02	-0.1	-0.03	-0.02	-0.01	-0.02	0	-0.03	0	-0.01	0	-0.01	0	-0.03	0.02	0.03	-0.07	0.02	0.03						
education=professional,course	0	-0.03	-0.07	-0.16	-0.13	-0.48	0.01	-0.08	0.01	-0.02	-0.03	0	-0.01	0	-0.01	-0.13	-0.21	-0.09	-0.01	-0.16	-0.27	-0.01	0.01	0	-0.03	-0.01	-0.02	0.07	0.01	0	0.01	-0.01	0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	-0.01	0.01	0	0.03	0.02	0.02					
education=university,degree	-0.04	-0.06	-0.18	0.3	-0.32	-0.04	-0.05	0.24	-0.02	0.09	0.05	0.01	-0.1	0.11	-0.01	-0.24	-0.39	-0.17	-0.3	-0.27	-0.01	-0.01	0.01	0.01	-0.12	-0.02	-0.04	0.14	0.02	0.05	0.02	0.04	0.01	0.04	-0.01	-0.01	0.01	-0.01	0.01	0	0.04	-0.05	-0.08	0.08	-0.04	-0.04					
education=illiterate	0.02	0	-0.01	-0.01	0.01	-0.01	0.01	-0.01	0	0.02	0.01	0	0.01	-0.01	0	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	1	0	0	-0.01	-0.01	0.01	0.01	0	0	0	0	0	0	0	-0.01	0	0	0	0	0	0	-0.01	0	0	0	0				
default=yes	0	0	0	-0.01	0	0.01	0	0	0.02	0	0	0	0.01	-0.01	0	0	0	0	0	0	0	0	0	0	0	-0.01	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0	0	0	0	0.01	0	0.01	0.01			
housing=yes	0	0	0	0.01	-0.01	0.01	0	-0.01	0.01	0	0	0	-0.01	0.01	0	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	0.01	0	0	0.04	1	0	-0.01	0.02	-0.05	0	0.03	0.01	0.03	0.01	0.01	-0.01	0.01	-0.01	0.02	-0.03	0.02	0.01	-0.06	-0.08	-0.03	-0.06	-0.05			
loan=yes	-0.01	0	0	0.02	-0.01	-0.01	-0.01	0	0	-0.01	-0.01	0	0	0	-0.01	0	0	-0.01	-0.01	0	0	0	0	0	0.04	1	0	-0.01	0.02	0	-0.01	0	0.01	0	0	0.01	-0.01	0	0	0	0	0	0	-0.01	-0.01	0	0	0			
month=may	-0.07	-0.02	0.06	-0.05	0.14	-0.06	-0.06	-0.01	-0.02	0	-0.02	0.02	-0.02	-0.02	0.02	0.04	0.06	0.09	-0.03	-0.12	-0.01	-0.01	-0.02	0	0	0	0	-0.27	-0.33	-0.3	-0.09	-0.24	-0.05	-0.08	-0.19	-0.08	-0.03	0.01	0.03	-0.04	0.03	-0.03	-0.01	-0.02	0.06	-0.07	-0.11	-0.06	-0.01	-0.13	-0.17
month=jun	-0.01	0	0.01	-0.01	0.03	-0.03	-0.02	-0.01	0.02	0.01	0.01	-0.01	0.02	-0.02	-0.01	0.02	0	0.01	-0.02	-0.01	0	-0.05	-0.01	-0.02	0	0	0	-0.27	1	-0.18	-0.16	-0.05	-0.13	-0.03	-0.04	-0.1	-0.05	0.03	-0.01	-0.03	0.03	0.07	-0.07	0.09	-0.09	0.14	0.16				
month=jul	-0.04	0.02	0.02	0	-0.02	-0.01	0	-0.01	-0.01	0	0	-0.03	0.01	-0.02	0.02	-0.03	0.01	0	-0.02	-0.03	-0.18	-0.01	0	0	0.02	-0.33	-0.18	1	-0.19	-0.16	-0.06	-0.15	-0.03	-0.05	-0.12	-0.05	0.01	0.02	0.01	-0.06	0.14	-0.13	0.25	-0.19	0.28	0.3					
month=aug	0.07	0.02	-0.07	0.08	-0.13	0.14	0.03	-0.02	-0.01	0	0	-0.03	-0.05	-0.01	0.04	-0.02	-0.06	-0.1	0	0.07	-0.03	0.01	0.02	0	-0.3	-0.16	-0.19	1	-0.06	-0.14	-0.03	-0.05	-0.11	-0.05	-0.01	0.01	0	-0.02	0.02	-0.05	0.08	-0.09	0.18	-0.19	0.45	-0.16	-0.19				
month=oct	0.05	0	-0.02	0.01	-0.05	0	0.09	0	0.01	0	-0.01																																								

**7. Cuál variable tiene la correlación más alta con la variable objetivo?**

nr.employed de 0.35468

## 8. Cuál variable tiene la correlación más baja con la variable objetivo?

default de 0.00304

```
Ranked attributes:
0.35468  17 nr.employed
0.30777  16 euribor3m
0.29833  13 emp.var.rate
0.23018  11 previous
0.18091  12 poutcome
0.13621  14 cons.price.idx
0.06636  10 campaign
0.05675   8 month
0.05488  15 cons.conf.idx
0.04248   3 marital
0.03611   2 job
0.0304    1 age
0.03035   4 education
0.01146   9 day_of_week
0.01109   6 housing
0.00447   7 loan
0.00304   5 default
```

Aunque se considera que las variables con una correlación menor a 0.3 se consideran irrelevantes, solo se decide eliminar las dos últimas variables ya que estas no ayudarían en prácticamente nada en un análisis y no se eliminan todas las que están por debajo del 0.3 para no quedar con tan pocas variables predictoras.

## 9, ¿Cuántos componentes se pueden seleccionar con PCA?

Se deben seleccionar el número de componentes con los cuales se logre asegurar como mínimo el 0.75 de la varianza de los datos, por lo que se decide realizar varios análisis con diferentes cantidades de variables en las ecuaciones.

Con ecuaciones de máximo 3 variables: se deben seleccionar 25 componentes

eigenvalue	proportion	cumulative	
2.72276	0.06051	0.06051	-0.366nr.employed+0.336marital=single-0.33cons.price.idx...
2.38484	0.053	0.1135	0.362age-0.355nr.employed-0.325marital=single...
2.23501	0.04967	0.16317	0.409month=aug+0.366cons.conf.idx-0.33job=blue-collar...
1.80508	0.04011	0.20328	0.559education=professional.course+0.551job=technician-0.348job=admin....
1.67857	0.0373	0.24058	0.554education=high.school+0.439job=services-0.314education=university.degree...
1.60848	0.03574	0.27633	-0.446cons.conf.idx-0.387month=may+0.363month=apr...
1.52243	0.03383	0.31016	0.401marital=married-0.346education=basic.4y-0.277job=retired...
1.37712	0.0306	0.34076	-0.535month=jun+0.292marital=divorced-0.289cons.price.idx...
1.3323	0.02961	0.37037	-0.455marital=divorced+0.433day_of_week=thu-0.37month=may...
1.28295	0.02851	0.39888	0.358day_of_week=mon-0.352day_of_week=tue+0.334day_of_week=fri...
1.26024	0.02801	0.42688	0.697day_of_week=mon-0.465day_of_week=fri-0.464day_of_week=thu...
1.2488	0.02775	0.45463	0.698day_of_week=wed-0.679day_of_week=tue-0.105month=mar...
1.19963	0.02666	0.48129	0.486day_of_week=thu-0.463day_of_week=fri+0.399month=may...
1.18417	0.02631	0.50761	0.328month=nov-0.311job=admin.-0.287day_of_week=wed...
1.16856	0.02597	0.53358	-0.477job=housemaid+0.424education=basic.9y-0.413education=basic.4y...
1.12548	0.02501	0.55859	-0.487month=jul+0.413month=jun-0.282day_of_week=fri...
1.08963	0.02421	0.5828	0.633education=basic.6y+0.307job=management-0.28job=self-employed...
1.07741	0.02394	0.60674	-0.362job=self-employed+0.306job=entrepreneur+0.274month=nov...
1.04751	0.02328	0.63002	-0.526job=entrepreneur+0.435job=unemployed+0.291job=management...
1.03315	0.02296	0.65298	0.551job=self-employed+0.378education=basic.6y-0.359month=oct...
1.02666	0.02281	0.67579	0.508month=mar-0.467month=sep-0.362month=apr...
1.02234	0.02272	0.69851	0.555loan=yes+0.386housing=yes+0.291job=entrepreneur...
1.01257	0.0225	0.72102	-0.446default=yes-0.382job=entrepreneur-0.367month=dec...
1.00539	0.02234	0.74336	-0.503month=oct-0.337job=unemployed-0.335loan=yes...
1.00243	0.02228	0.76563	0.712month=dec-0.358month=mar-0.258job=entrepreneur...
0.99259	0.02206	0.78769	-0.83education=illiterate-0.268default=yes+0.217job=self-employed...
0.98754	0.02195	0.80964	0.698default=yes-0.442job=unemployed-0.357education=illiterate...
0.96786	0.02151	0.83114	-0.432job=housemaid-0.419month=mar+0.3 job=retired...

Con ecuaciones de máximo 4 variables: se deben seleccionar 25 componentes

eigenvalue	proportion	cumulative	
2.72276	0.06051	0.06051	-0.366nr.employed+0.336marital=single-0.33cons.price.idx-0.321marital=married...
2.38484	0.053	0.1135	0.362age-0.355nr.employed-0.325marital=single+0.308marital=married...
2.23501	0.04967	0.16317	0.409month=aug+0.366cons.conf.idx-0.33job=blue-collar+0.315education=university.degree...
1.80508	0.04011	0.20328	0.559education=professional.course+0.551job=technician-0.348job=admin.-0.332education=university.degree...
1.67857	0.0373	0.24058	0.554education=high.school+0.439job=services-0.314education=university.degree-0.294job=blue-collar...
1.60848	0.03574	0.27633	-0.446cons.conf.idx-0.387month=may+0.363month=apr+0.247month=jul...
1.52243	0.03383	0.31016	0.401marital=married-0.346education=basic.4y-0.277job=retired-0.272marital=single...
1.37712	0.0306	0.34076	-0.535month=jun+0.292marital=divorced-0.289cons.price.idx-0.287day_of_week=fri...
1.3323	0.02961	0.37037	-0.455marital=divorced+0.433day_of_week=thu-0.37month=may+0.236month=aug...
1.28295	0.02851	0.39888	0.358day_of_week=mon-0.352day_of_week=tue+0.334day_of_week=fri-0.318day_of_week=wed...
1.26024	0.02801	0.42688	0.697day_of_week=mon-0.465day_of_week=fri-0.464day_of_week=thu+0.134day_of_week=tue...
1.2488	0.02775	0.45463	0.698day_of_week=wed-0.679day_of_week=tue-0.105month=mar+0.075month=nov...
1.19963	0.02666	0.48129	0.486day_of_week=thu-0.463day_of_week=fri+0.399month=may+0.351month=nov...
1.18417	0.02631	0.50761	0.328month=nov-0.311job=admin.-0.287day_of_week=wed-0.272month=jul...
1.16856	0.02597	0.53358	-0.477job=housemaid+0.424education=basic.9y-0.413education=basic.4y+0.307job=admin...
1.12548	0.02501	0.55859	-0.487month=jul+0.413month=jun-0.282day_of_week=fri+0.28 job=admin...
1.08963	0.02421	0.5828	0.633education=basic.6y+0.307job=management-0.28job=self-employed-0.257job=entrepreneur...
1.07741	0.02394	0.60674	-0.362job=self-employed+0.306job=entrepreneur+0.274month=nov+0.267education=basic.6y...
1.04751	0.02328	0.63002	-0.526job=entrepreneur+0.435job=unemployed+0.291job=management-0.284job=self-employed...
1.03315	0.02296	0.65298	0.551job=self-employed+0.378education=basic.6y-0.359month=oct+0.287month=sep...
1.02666	0.02281	0.67579	0.508month=mar-0.467month=sep-0.362month=apr+0.258month=nov...
1.02234	0.02272	0.69851	0.555loan=yes+0.386housing=yes+0.291job=entrepreneur+0.279month=dec...
1.01257	0.0225	0.72102	-0.446default=yes-0.382job=entrepreneur-0.367month=dec-0.355job=unemployed...
1.00539	0.02234	0.74336	-0.503month=oct-0.337job=unemployed-0.335loan=yes-0.302housing=yes...
1.00243	0.02228	0.76563	0.712month=dec-0.358month=mar-0.258job=entrepreneur-0.22month=sep...
0.99259	0.02206	0.78769	-0.83education=illiterate-0.268default=yes+0.217job=self-employed+0.167job=entrepreneur...
0.98754	0.02195	0.80964	0.698default=yes-0.442job=unemployed-0.357education=illiterate+0.198loan=yes...
0.96786	0.02151	0.83114	-0.432job=housemaid-0.419month=mar+0.3 job=retired+0.285job=unemployed...

Con ecuaciones de máximo 6 variables: se deben seleccionar 25 componentes

eigenvalue	proportion	cumulative	
2.72276	0.06051	0.06051	-0.366nr.employed+0.336marital=single-0.33cons.price.idx-0.321marital=married-0.283job=blue-collar+0.25 previous...
2.38484	0.053	0.1135	0.362age-0.355nr.employed-0.325marital=single+0.308marital=married-0.303cons.price.idx+0.279job=retired...
2.23501	0.04967	0.16317	0.409month=aug+0.366cons.conf.idx-0.33job=blue-collar+0.315education=university.degree-0.261month=may-0.253education=basi
1.80508	0.04011	0.20328	0.559education=professional.course+0.551job=technician-0.348job=admin.-0.332education=university.degree-0.19job=managemen
1.67857	0.0373	0.24058	0.554education=high.school+0.439job=services-0.314education=university.degree-0.294job=blue-collar-0.267education=basic.9
1.60848	0.03574	0.27633	-0.446cons.conf.idx-0.387month=may+0.363month=apr+0.247month=jul-0.243month=aug+0.188education=professional.course...
1.52243	0.03383	0.31016	0.401marital=married-0.346education=basic.4y-0.277job=retired-0.272marital=single-0.234marital=divorced+0.228job=services
1.37712	0.0306	0.34076	-0.535month=jun+0.292marital=divorced-0.289cons.price.idx-0.287day_of_week=fri+0.247month=jul-0.213month=sep...
1.3323	0.02961	0.37037	-0.455marital=divorced+0.433day_of_week=thu-0.37month=may+0.236month=aug+0.226marital=married-0.223day_of_week=fri...
1.28295	0.02851	0.39888	0.358day_of_week=mon-0.352day_of_week=tue+0.334day_of_week=fri-0.318day_of_week=wed-0.291month=nov+0.237month=aug...
1.26024	0.02801	0.42688	0.697day_of_week=mon-0.465day_of_week=fri-0.464day_of_week=thu+0.134day_of_week=tue-0.1marital=divorced+0.092month=mar...
1.2488	0.02775	0.45463	0.698day_of_week=wed-0.679day_of_week=tue-0.105month=mar+0.075month=nov-0.071month=may+0.056marital=divorced...
1.19963	0.02666	0.48129	0.486day_of_week=thu-0.463day_of_week=fri+0.399month=may-0.351month=nov-0.17month=aug+0.144cons.price.idx...
1.18417	0.02631	0.50761	0.328month=nov-0.311job=admin.-0.287day_of_week=wed-0.272month=jul+0.269education=basic.9y+0.243day_of_week=mon...
1.16856	0.02597	0.53358	-0.47job=housemaid+0.424education=basic.9y-0.413education=basic.4y+0.307job=admin.-0.265job=management-0.187job=student...
1.12548	0.02501	0.55859	-0.487month=jul+0.413month=jun-0.282day_of_week=fri+0.28 job=admin.-0.261job=management-0.228campaign...
1.08963	0.02421	0.5828	0.633education=basic.6y+0.307job=management-0.28job=self-employed-0.257job=entrepreneur-0.23job=housemaid-0.205month=may...
1.07741	0.02394	0.60674	-0.362job=self-employed+0.306job=entrepreneur+0.274month=nov+0.267education=basic.6y-0.259month=apr+0.207job=housemaid...
1.04751	0.02328	0.63002	-0.526job=entrepreneur+0.435job=unemployed+0.291job=management-0.284job=self-employed-0.276education=illiterate-0.239mont
1.03315	0.02296	0.65298	0.551job=self-employed+0.378education=basic.6y-0.359month=oct+0.287month=sep+0.239housing=yes-0.233job=management...
1.02666	0.02281	0.67579	0.508month=mar-0.467month=sep-0.362month=apr+0.258month=nov+0.245month=oct-0.167marital=divorced...
1.02234	0.02272	0.69851	0.555loan=yes+0.386housing=yes+0.291job=entrepreneur+0.279month=dec-0.253default=yes-0.221job=unemployed...
1.01257	0.0225	0.72102	-0.446default=yes-0.382job=entrepreneur-0.367month=dec-0.355job=unemployed+0.249loan=yes+0.189month=oct...
1.00539	0.02234	0.74336	-0.503month=oct-0.337job=unemployed-0.335loan=yes-0.302housing=yes-0.299default=yes+0.261month=mar...
1.00243	0.02228	0.76563	0.712month=dec-0.358month=mar-0.258job=entrepreneur-0.22month=sep+0.203job=self-employed+0.186job=housemaid...
0.99259	0.02206	0.78769	-0.83education=illiterate-0.268default=yes+0.217job=self-employed+0.167job=entrepreneur-0.153loan=yes+0.152education=basi
0.98754	0.02195	0.80964	0.698default=yes-0.442job=unemployed-0.357education=illiterate+0.198loan=yes+0.157job=self-employed+0.15 month=nov...
0.96786	0.02151	0.83114	-0.432job=housemaid-0.419month=mar+0.3 job=retired+0.285job=unemployed-0.252marital=divorced+0.222marital=single...

En consecuencia, es posible considerar quedarse con los (mínimo) 25 componentes para reemplazar las variables actuales, ya que en realidad se tendría que trabajar con 45 variables considerando que se deben realizar las dummies a todas las categorías.

**10. Cuánto es el porcentaje en el balanceo?**

% Balanceo	% Aumento	Categoría	Evidencia												
100	687.67	yes	<div> <div> <div> <div>Name: y</div> <div>Missing: 0 (0%)</div> </div> <div> <div>Distinct: 2</div> </div> <div> <div>Type: Nominal</div> <div>Unique: 0 (0%)</div> </div> </div> <table border="1"> <thead> <tr> <th>No.</th> <th>Label</th> <th>Count</th> <th>Weight</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>no</td> <td>36548</td> <td>36548.0</td> </tr> <tr> <td>2</td> <td>yes</td> <td>36547</td> <td>36547.0</td> </tr> </tbody> </table> <div> <div> <div> <div>36548</div> <div></div> </div> <div> <div>36547</div> <div></div> </div> </div> <div> <div> <div>Class: y (Nom)</div> <div>Visualize All</div> </div> </div> </div> </div>	No.	Label	Count	Weight	1	no	36548	36548.0	2	yes	36547	36547.0
No.	Label	Count	Weight												
1	no	36548	36548.0												
2	yes	36547	36547.0												
75	490.75	yes	<div> <div> <div> <div>Name: y</div> <div>Missing: 0 (0%)</div> </div> <div> <div>Distinct: 2</div> </div> <div> <div>Type: Nominal</div> <div>Unique: 0 (0%)</div> </div> </div> <table border="1"> <thead> <tr> <th>No.</th> <th>Label</th> <th>Count</th> <th>Weight</th> </tr> </thead> <tbody> <tr> <td>1</td> <td>no</td> <td>36548</td> <td>36548.0</td> </tr> <tr> <td>2</td> <td>yes</td> <td>27410</td> <td>27410.0</td> </tr> </tbody> </table> <div> <div> <div> <div>36548</div> <div></div> </div> <div> <div>27410</div> <div></div> </div> </div> <div> <div> <div>Class: y (Nom)</div> <div>Visualize All</div> </div> </div> </div> </div>	No.	Label	Count	Weight	1	no	36548	36548.0	2	yes	27410	27410.0
No.	Label	Count	Weight												
1	no	36548	36548.0												
2	yes	27410	27410.0												

50

293.83

yes

Name: y		Type: Nominal	
Missing: 0 (0%)		Distinct: 2	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	no	36548	36548.0
2	yes	18273	18273.0

Bar chart showing the distribution of the nominal variable 'y'. The 'no' category (blue bar) has a count of 36548, and the 'yes' category (red bar) has a count of 18273. The chart includes a legend and a 'Visualize All' button.

**11. ¿Si el método que deseamos aplicar es para variables categóricas, cuáles variables se deben transformar? ¿Cómo?**

age, campaign, previous, cons.price.idx, cons.conf.idx y nr.employed, a la cuales se le debe emplear el método de discretización que se encarga de crear rangos de los datos numéricos que se convierten en categorías.