

Catedra 1

Raúl Urzua

Institución Académica: Universidad de las Américas.

Programa: Data Science.

Asignatura: Programación en R.

Alumno: Raúl Urzua.

1. Carga de datos y resumen básico

1.1. El problema del negocio

Una entidad bancaria contrata a una empresa de marketing, encargada de contactar telefónicamente a posibles clientes, para determinar si están interesados o en adquirir una cuenta de depósito con el banco.

1.2. Objetivo Principal

Realizar un Análisis exploratorio de datos (EDA).

1.3. Objetivos Específicos

- Generas preguntas acerca de tus datos.
- Buscas respuestas visualizando, transformando y modelando tus datos.
- Usas lo que has aprendido para refinar tus preguntas y/o generar nuevas interrogantes.

1.4. Dataset

La información recolectada por la empresa de mercadeo se encuentra en un archivo CSV (dataset_banco_clean.csv) con 45189 filas y 17 columnas.

1.4.1. Variables

Las columnas son:

- age: edad. Tipo de variable: variable numérica.
- job: tipo de trabajo. Tipo de variable: variable categórica (admin, unknown, unemployed, management, housemaid, entrepreneur, student, blue-collar, self-employed, retired, technician, services).
- marital: estado civil. Tipo de variable: variable categórica (married, divorced, single).
- education: nivel educativo. Tipo de variable: variable categórica (unknown, primary, secondary, tertiary).
- default: si dejó de pagar sus obligaciones. Tipo de variable: variable categórica (yes, no).
- balance: saldo promedio anual en euros. Tipo de variable: variable numérica.
- housing: ¿tiene o no crédito hipotecario? Tipo de variable: variable categórica (yes, no).
- loan: ¿tiene créditos de consumo? Tipo de variable: variable categórica (yes, no).
- contact: medio a través del cual fue contactado. Tipo de variable: variable categórica (unknown, telephone, cellular).
- day: último día del mes en el que fue contactada. Tipo de variable: variable numérica.
- month: último mes en el que fue contactada. Tipo de variable: variable categórica (jan, feb, mar, apr, may, jun, jul, aug, sep, oct, nov, dec).
- duration: duración (en segundos) del último contacto. Tipo de variable: variable numérica.
- campaign: número total de veces que fue contactada durante la campaña. Tipo de variable: variable numérica.
- pdays: número de días transcurridos después de haber sido contactado antes de la campaña actual. Tipo de variable: variable numérica (-1 indica que no fue contactado previamente).
- previous: número de veces que ha sido contactada antes de esta campaña. Tipo de variable: variable numérica.
- poutcome: resultado de la campaña de marketing anterior. Tipo de variable: variable categórica (unknown, other, failure, success).
- y: ¿el cliente se suscribió a un depósito al término?. Tipo de variable: variable categórica (yes, no).

1.4.2. Archivo del dataset

El archivo del dataset que se analizará, es “dataset_banco_clean.csv”.

```
#Se guardo un archivo formato "csv".  
#Corresponde a la base de datos que voy analizar.  
dataset_banco_clean <- read.csv("dataset_banco_clean.csv")
```

1.4.3. Encabezado del dataset

Se mostrara el encabezado del dataset.

```
#Se muestra el encabezado base de datos.  
head (dataset_banco_clean)
```

	X	age	job	marital	education	default	balance	housing	loan	contact	day
1	1	58	management	married	tertiary	no	2143	yes	no	unknown	5
2	2	44	technician	single	secondary	no	29	yes	no	unknown	5
3	3	33	entrepreneur	married	secondary	no	2	yes	yes	unknown	5
4	4	47	blue-collar	married	unknown	no	1506	yes	no	unknown	5
5	5	33	unknown	single	unknown	no	1	no	no	unknown	5
6	6	35	management	married	tertiary	no	231	yes	no	unknown	5

	month	duration	campaign	pdays	previous	poutcome	y	y_numerico
1	may	261	1	-1	0	unknown	no	0
2	may	151	1	-1	0	unknown	no	0
3	may	76	1	-1	0	unknown	no	0
4	may	92	1	-1	0	unknown	no	0
5	may	198	1	-1	0	unknown	no	0
6	may	139	1	-1	0	unknown	no	0

	education_numerico	job_numerico
1	3	2143
2	2	29
3	2	2
4	0	1506
5	0	1
6	3	231

1.4.4. Mediciones básicas del dataset

Se utilizara función summary(). el cual muestra un resumen de mediciones básicas, tales como: el mínimo, el máximo, la media, la mediana y los cuartiles 1 y 3 de un vector numérico .

```
#funcion summary
summary(dataset_banco_clean)
```

X	age	job	marital
Min. : 1	Min. :18.00	Length:45189	Length:45189
1st Qu.:11298	1st Qu.:33.00	Class :character	Class :character
Median :22595	Median :39.00	Mode :character	Mode :character
Mean :22595	Mean :40.94		
3rd Qu.:33892	3rd Qu.:48.00		
Max. :45189	Max. :95.00		
education	default	balance	housing
Length:45189	Length:45189	Min. : -8019	Length:45189
Class :character	Class :character	1st Qu.: 72	Class :character
Mode :character	Mode :character	Median : 448	Mode :character
		Mean : 1374	
		3rd Qu.: 1428	
		Max. :527532	
loan	contact	day	month
Length:45189	Length:45189	Min. : 1.00	Length:45189
Class :character	Class :character	1st Qu.: 8.00	Class :character
Mode :character	Mode :character	Median :16.00	Mode :character
		Mean :15.81	
		3rd Qu.:21.00	
		Max. :31.00	
duration	campaign	pdays	previous
Min. : 1.0	Min. : 1.000	Min. : -1.00	Min. : 0.0000
1st Qu.: 103.0	1st Qu.: 1.000	1st Qu.: -1.00	1st Qu.: 0.0000
Median : 180.0	Median : 2.000	Median : -1.00	Median : 0.0000
Mean : 258.1	Mean : 2.763	Mean : 40.18	Mean : 0.5742
3rd Qu.: 319.0	3rd Qu.: 3.000	3rd Qu.: -1.00	3rd Qu.: 0.0000
Max. :4918.0	Max. :63.000	Max. :871.00	Max. :58.0000
poutcome	y	y_numerico	education_numerico
Length:45189	Length:45189	Min. :0.000	Min. :0.000
Class :character	Class :character	1st Qu.:0.000	1st Qu.:2.000
Mode :character	Mode :character	Median :0.000	Median :2.000
		Mean :0.117	Mean :2.061
		3rd Qu.:0.000	3rd Qu.:3.000
		Max. :1.000	Max. :3.000
job_numerico			
Min. : -8019			
1st Qu.: 72			
Median : 448			

```
Mean    : 1374
3rd Qu.: 1428
Max.    :527532
```

2. Gráfico exploratorio principal

2.1. Gráfico de la variable “y”

El cual responde la siguiente pregunta, ¿el cliente se suscribió a un depósito al término?

```
#Instalar libreria tidyverse
install.packages("tidyverse")
```

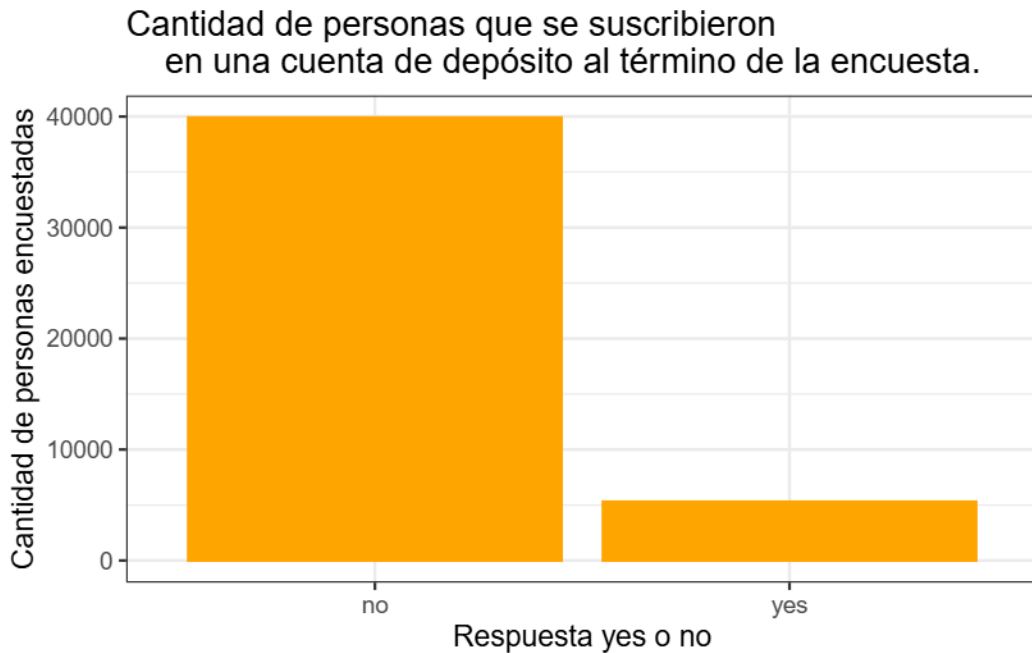
Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.4'
(as 'lib' is unspecified)

```
#Libreria tidyverse
library(tidyverse)
```

```
-- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
v dplyr      1.1.4      v readr      2.1.5
v forcats    1.0.0      v stringr    1.5.1
v ggplot2    3.5.2      v tibble     3.2.1
v lubridate  1.9.4      v tidyr      1.3.1
v purrr      1.0.4
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become
```

```
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = y),
    fill = "orange",
    color = "orange"
  ) +
  labs(
    title = "Cantidad de personas que se suscribieron
en una cuenta de depósito al término de la encuesta.",
    y = "Cantidad de personas encuestadas",
```

```
x = "Respuesta yes o no"
) +
theme_bw()
```



Claramente se ve en el gráfico, que la respuesta predominante es “no”.

Analizaremos los datos, para encontrar el porqué de esta situación.

3. Análisis

A continuación se analizarán una serie de gráficos realizados.

3.1. Análisis variable “duration” con respecto a la variable “y”

Duración de la llamada, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta

```
#Gráfico de la variable "duration"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = duration),
    fill = "blue",
```

```

    color = "blue"
  ) +
  facet_wrap(~y, nrow = 1) +
  labs(
    title = "Duración de la llamada, con respecto a las personas
            que se suscribieron a una cuenta de depósito al
            término de la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Duración de la llamada en segundos"
  ) +
  theme_bw()

```

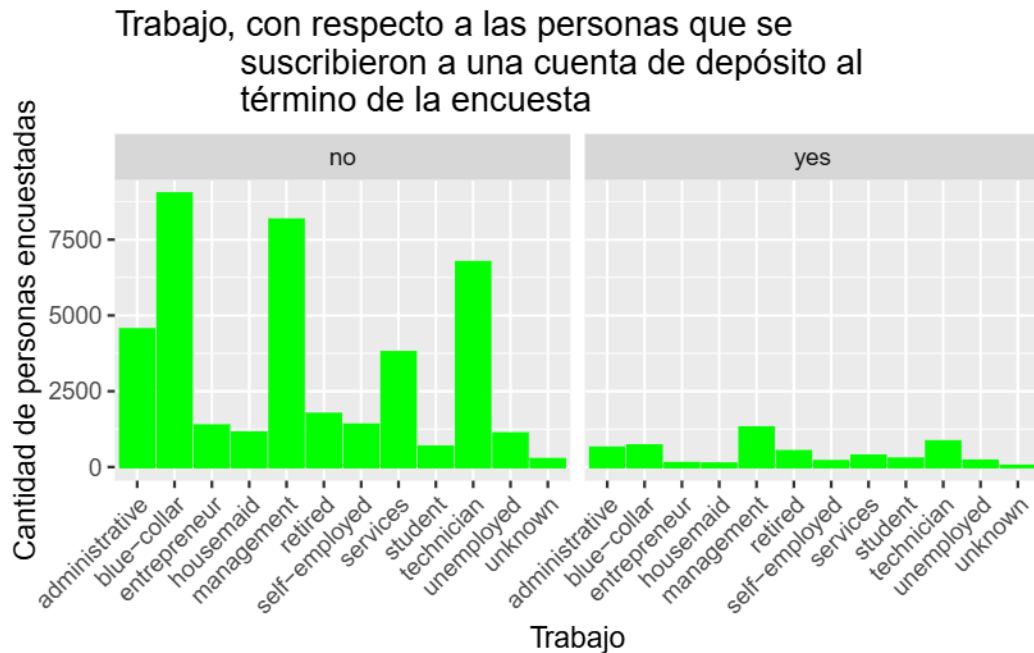


En una primera instancia pensé que la duración llamadas más largas, de los posibles clientes iban acceder positivamente a una suscripción de una cuenta en el banco, pero al ver el gráfico esto no se reflejado.

3.2. Análisis variable “job” con respecto a la variable “y”

Trabajo, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta.

```
#Gráfico de la variable "job"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = job),
    fill = "green",
    color = "green"
  ) +
  facet_wrap(~y, nrow = 1) +
  labs(
    title = "Trabajo, con respecto a las personas que se
            suscribieron a una cuenta de depósito al
            término de la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Trabajo"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```

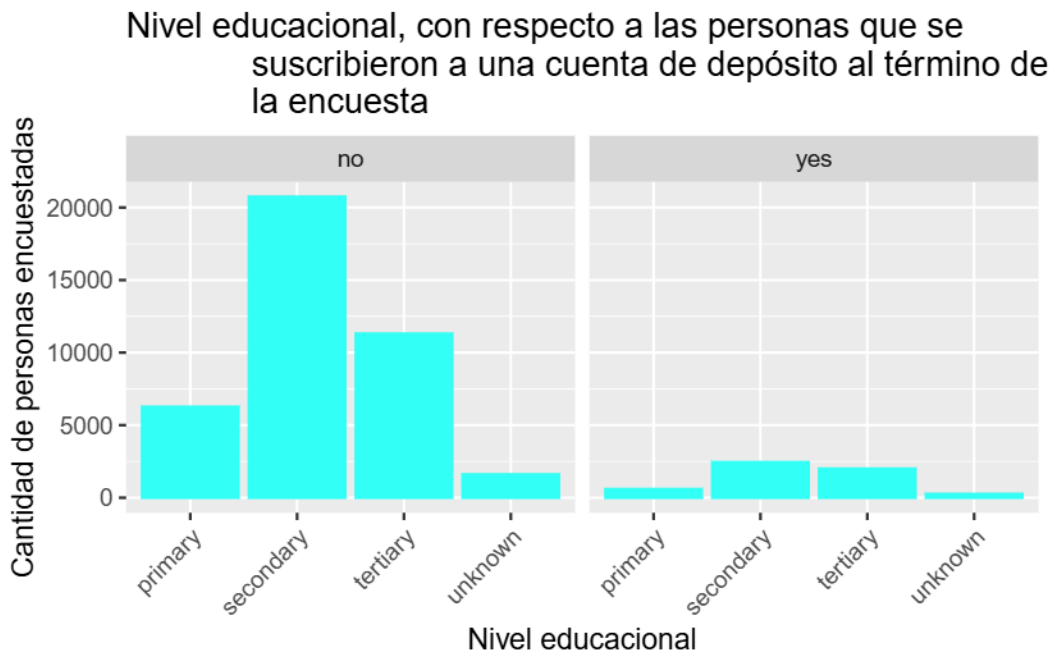


No distingo una relación clara entre el trabajo realizado con respecto a la respuesta de la encuesta.

3.3. Análisis variable “education” con respecto a la variable “y”

Nivel educacional, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta

```
#Gráfico de la variable "education"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = education),
    fill = "#33ffff",
    color = "#33ffff"
  ) +
  facet_wrap(~y, nrow = 1) +
  labs(
    title = "Nivel educacional, con respecto a las personas que se
            suscribieron a una cuenta de depósito al término de
            la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Nivel educacional"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



No distingo una relación clara entre el nivel educacional con respecto a la respuesta de la encuesta.

3.4. Análisis variable “balance” con respecto a la variable “y”

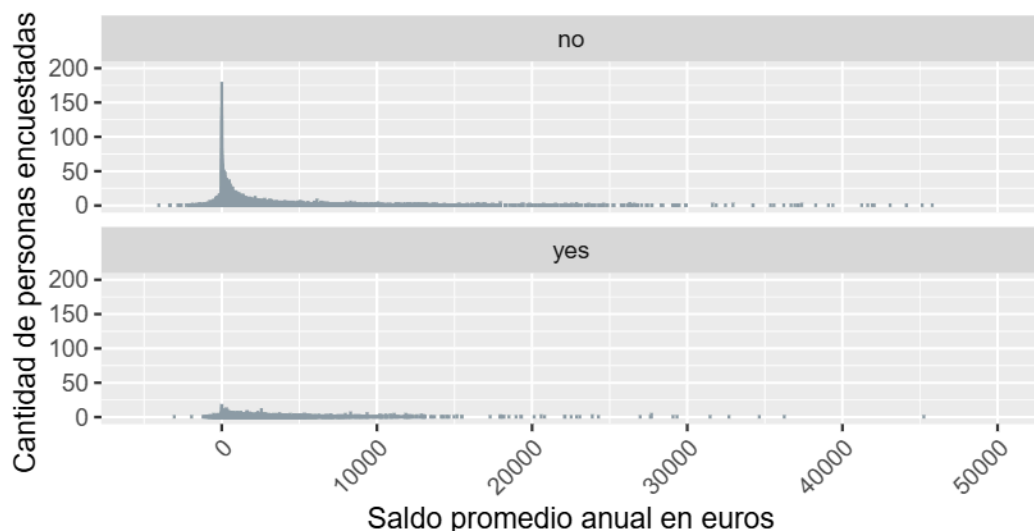
Saldo promedio anual en euros, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta.

```
#Gráfico de la variable "balance"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = balance),
    fill = "#8d9ca5",
    color = "#8d9ca5"
  ) +
  facet_wrap(~y, nrow = 2) +
  labs(
    title = "Saldo promedio anual en euros, con respecto a las
            personas que se suscribieron a una cuenta de depósito
            al término de la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Saldo promedio anual en euros"
  ) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  scale_x_continuous(limits = c(-5000, 50000)) +
  scale_y_continuous(limits = c(0, 200))
```

Warning: Removed 21 rows containing non-finite outside the scale range
(`stat_count()`).

Warning: Removed 2 rows containing missing values or values outside the scale range
(`geom_bar()`).

Saldo promedio anual en euros, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta



No distingo una relación clara entre el saldo de promedio anual con respecto a la respuesta de la encuesta.

3.5. Análisis variable “previous” con respecto a la variable “y”

Número de veces que ha sido contactada antes de esta campaña, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta.

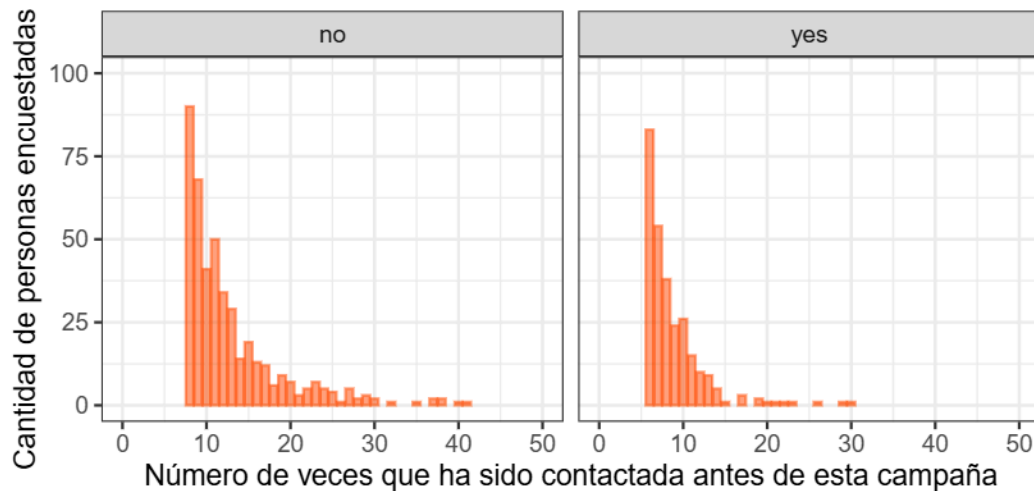
```
#Gráfico de la variable "previous"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = previous),
    fill = "#FF4B0082",
    color = "#FF4B0082"
  ) +
  facet_wrap(~y, nrow = 1) +
  labs(
    title = "Número de veces que ha sido contactada antes de
            esta campaña, con respecto a las personas que se
            suscribieron a una cuenta de depósito al término
            de la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Número de veces que ha sido contactada antes de esta campaña"
```

```
) +
scale_x_continuous(limits = c(0, 50)) +
scale_y_continuous(limits = c(0, 100)) +
theme_bw()
```

Warning: Removed 3 rows containing non-finite outside the scale range
(`stat_count()`).

Warning: Removed 14 rows containing missing values or values outside the scale range
(`geom_bar()`).

Número de veces que ha sido contactada antes de esta campaña, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta



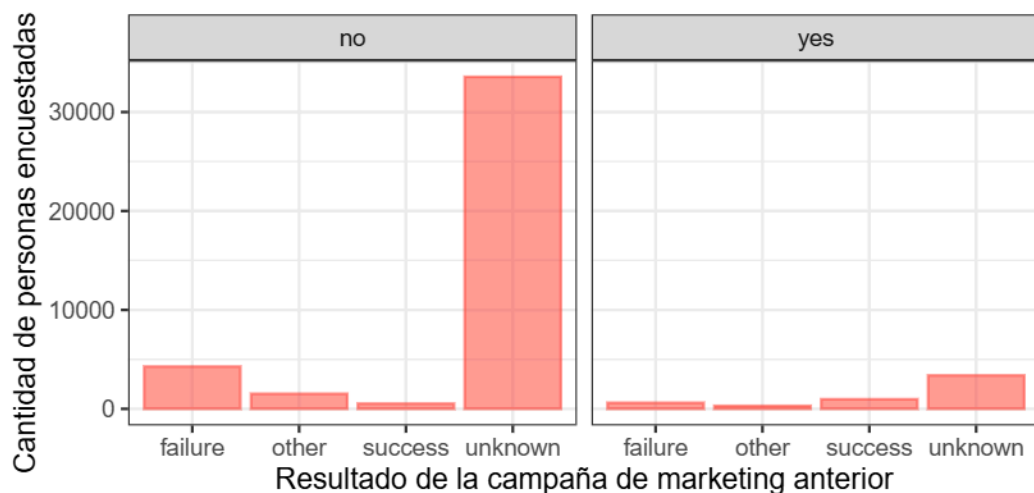
La gráfica es bastante similar entre la respuesta “sí” y “no”, se puede decir hasta cierto punto que el número de veces que han sido contactada antes de esta campaña no es relevante para su respuesta.

3.5. Análisis variable “poutcome” con respecto a la variable “y”

Resultado de la campaña de marketing anterior, con respecto a las personas que se suscribieron a una cuenta de depósito al término de la encuesta.

```
#Gráfico de la variable "poutcome"
ggplot(data = dataset_banco_clean) +
  geom_bar(
    mapping = aes(x = poutcome),
    fill = "#FF483D8B",
    color = "#FF483D8B"
  ) +
  facet_wrap(~y, nrow = 1) +
  labs(
    title = "Resultado de la campaña de marketing anterior,
    con respecto a las personas que se
    suscribieron a una cuenta de deposito al termino
    de la encuesta",
    y = "Cantidad de personas encuestadas",
    x = "Resultado de la campaña de marketing anterior"
  ) +
  theme_bw()
```

Resultado de la campaña de marketing anterior,
con respecto a las personas que se
suscribieron a una cuenta de deposito al termino
de la encuesta



Se observa en la gráfica que la mayoría de los personas encuestadas desconocía la campaña de marketing anterior, tanto en las personas que si se suscribieron como en las que no.

4. Agregar columnas

4.1. Columna y_numerico

Se transformó la variable categórica “y”, ha una variable numérica “y_numerico”. Y pare ello se agregó la columna.

```
#funcion crear columna nueva y_numerico
#convertir variable categorica a numerica
cont0<-0; #contador 0
cont1<-0; #contador 1
#Agregar columna
dataset_banco_clean$y_numerico <- dataset_banco_clean$balance
print("Se agrego columna")
```

```
[1] "Se agrego columna"
```

```
#Guardar dataset
write.csv(dataset_banco_clean, file = "dataset_banco_clean.csv")
print("Se guardo dataset")
```

```
[1] "Se guardo dataset"
```

```
for (i in 1:45189) {
  if (dataset_banco_clean$y[i] == "no"){
    dataset_banco_clean$y_numerico[i] <- as.numeric(0)
    cont0<-cont0+1;
  }
  if (dataset_banco_clean$y[i] == "yes"){
    dataset_banco_clean$y_numerico[i] <- as.numeric(1)
    cont1<-cont1+1;
  }
}
#Imprime la cantidad de "no"
print(paste("cantidad de no: ", cont0))
```

```
[1] "cantidad de no: 39904"
```

```
#Imprime la cantidad de "yes"
print(paste("cantidad de yes: ", cont1))
```

```
[1] "cantidad de yes: 5285"
```

```
#Cabecera
print("Variable job_numerico")
```

```
[1] "Variable job_numerico"
```

```
head(dataset_banco_clean$y_numerico)
```

```
[1] 0 0 0 0 0 0
```

4.2. Columna education_numerico

Se transformó la variable categórica “education”, ha una variable numérica “education_numerico”. Y pare ello se agregó la columna.

```
#funcion crear columna nueva education_numerico
#convertir variable categorica a numerica
cont0<-0; #contador 0
cont1<-0; #contador 1
cont2<-0; #contador 2
cont3<-0; #contador 3
#Agregar columna
dataset_banco_clean$education_numerico <- dataset_banco_clean$balance
print("Se agrego columna")
```

```
[1] "Se agrego columna"
```

```
#Guardar dataset
write.csv(dataset_banco_clean, file = "dataset_banco_clean.csv")
print("Se guardo dataset")
```

```
[1] "Se guardo dataset"
```

```

for (i in 1:45189) {
  if (dataset_banco_clean$education[i] == "unknown"){
    dataset_banco_clean$education_numerico[i] <- as.numeric(0);
    cont0<-cont0+1;
  }
  if (dataset_banco_clean$education[i] == "primary"){
    dataset_banco_clean$education_numerico[i] <- as.numeric(1);
    cont1<-cont1+1;
  }
  if (dataset_banco_clean$education[i] == "secondary"){
    dataset_banco_clean$education_numerico[i] <- as.numeric(2);
    cont2<-cont2+1;
  }
  if (dataset_banco_clean$education[i] == "tertiary"){
    dataset_banco_clean$education_numerico[i] <- as.numeric(3);
    cont3<-cont3+1;
  }
}
print(paste("cantidad de unknown: ", cont0))

```

```
[1] "cantidad de unknown: 1857"
```

```
print(paste("cantidad de primary: ", cont1))
```

```
[1] "cantidad de primary: 6847"
```

```
print(paste("cantidad de secondary: ", cont2))
```

```
[1] "cantidad de secondary: 23187"
```

```
print(paste("cantidad de tertiary: ", cont3))
```

```
[1] "cantidad de tertiary: 13298"
```

```

#Cabecera
print("Variable job_numerico")

```

```
[1] "Variable job_numerico"
```



```
head(dataset_banco_clean$education_numerico)
```

```
[1] 3 2 2 0 0 3
```

4.2. Columna job_numerico

Se transformó la variable categórica “job”, ha una variable numérica “job_numerico”. Y pare ello se agregó la columna.

```
#convertir variable categorica a numerica
cont0<-0; #contador 0
cont1<-0; #contador 1
cont2<-0; #contador 2
cont3<-0; #contador 3
cont4<-0; #contador 4
cont5<-0; #contador 5
cont6<-0; #contador 6
cont7<-0; #contador 7
cont8<-0; #contador 8
cont9<-0; #contador 9
cont10<-0; #contador 10
cont11<-0; #contador 11
#Agregar columna
dataset_banco_clean$job_numerico <- dataset_banco_clean$balance
print("Se agrego columna")
```

```
[1] "Se agrego columna"
```

```
#Guardar dataset
write.csv(dataset_banco_clean, file = "dataset_banco_clean.csv")
print("Se guardo dataset")
```

```
[1] "Se guardo dataset"
```

```
for (i in 1:45189) {
  if (dataset_banco_clean$job[i] == "unknown"){
    dataset_banco_clean$job_numerico[i] <- as.numeric(0);
    cont0<-cont0+1;
  }
}
```

```

if (dataset_banco_clean$job[i] == "unemployed"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(1);
  cont1<-cont1+1;
}
if (dataset_banco_clean$job[i] == "student"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(2);
  cont2<-cont2+1;
}
if (dataset_banco_clean$job[i] == "retired"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(3);
  cont3<-cont3+1;
}
if (dataset_banco_clean$job[i] == "housemaid"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(4);
  cont4<-cont4+1;
}
if (dataset_banco_clean$job[i] == "self-employed"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(5);
  cont5<-cont5+1;
}
if (dataset_banco_clean$job[i] == "services"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(6);
  cont6<-cont6+1;
}
if (dataset_banco_clean$job[i] == "technician"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(7);
  cont7<-cont7+1;
}
if (dataset_banco_clean$job[i] == "administrative"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(8);
  cont8<-cont8+1;
}
if (dataset_banco_clean$job[i] == "blue-collar"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(9);
  cont9<-cont9+1;
}
if (dataset_banco_clean$job[i] == "entrepreneur"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(10);
  cont10<-cont10+1;
}
if (dataset_banco_clean$job[i] == "management"){
  dataset_banco_clean$job_numerico[i] <- as.numeric(11);
}

```

```
    cont11<-cont11+1;
  }
}
print(paste("Cantidad de unknown: ", cont0))
```

```
[1] "Cantidad de unknown: 288"
```

```
print(paste("Cantidad de unemployed: ", cont1))
```

```
[1] "Cantidad de unemployed: 1303"
```

```
print(paste("Cantidad de student: ", cont2))
```

```
[1] "Cantidad de student: 937"
```

```
print(paste("Cantidad de housemaid: ", cont4))
```

```
[1] "Cantidad de housemaid: 1240"
```

```
print(paste("Cantidad de self-employed: ", cont5))
```

```
[1] "Cantidad de self-employed: 1578"
```

```
print(paste("Cantidad de services: ", cont6))
```

```
[1] "Cantidad de services: 4152"
```

```
print(paste("Cantidad de technician: ", cont7))
```

```
[1] "Cantidad de technician: 7592"
```

```
print(paste("Cantidad de administrative: ", cont8))
```

```
[1] "Cantidad de administrative: 5168"
```

```
print(paste("Cantidad de blue-collar: ", cont9))
```

```
[1] "Cantidad de blue-collar: 9727"
```

```
print(paste("Cantidad de entrepreneur: ", cont10))
```

```
[1] "Cantidad de entrepreneur: 1486"
```

```
print(paste("Cantidad de management: ", cont11))
```

```
[1] "Cantidad de management: 9455"
```

```
#Cabecera  
print("Variable job_numerico")
```

```
[1] "Variable job_numerico"
```

```
head(dataset_banco_clean$job_numerico)
```

```
[1] 11 7 10 9 0 11
```

5. Funcion

5.1 Entre job_numerico y balance

Se realizo una función entre el trabajo y saldo en euros de los usuarios encuestados..

Voy analizar estás dos variables por pienso que el saldo en euros y el trabajo de los usuarios deben tener una relación, en el momento de suscribirse a una cuenta en el banco.

Porque una persona con mejor trabajo debe tener un mayor saldo en su cuenta y por ende debe querer mas productos financieros, como una suscripcion a una cuenta en el banco.

```

#Grafico muestra el resultado de la campaña de marketing anterior, con respecto a las personas
ggplot(data = dataset_banco_clean) +
  geom_point(mapping = aes(x = dataset_banco_clean$job_numerico,
                           y = dataset_banco_clean$balance)) +

  facet_grid(~dataset_banco_clean$y) +
  scale_y_continuous(limits = c(0, 100000)) +
  theme_bw() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Resultado de la campaña de marketing anterior,
            con respecto a las personas que se
            suscribieron a una cuenta de deposito al termino
            de la encuesta",
    y = "Saldo promedio anual en euros (balance)",
    x = "Trabajo (job_numero)"
  )

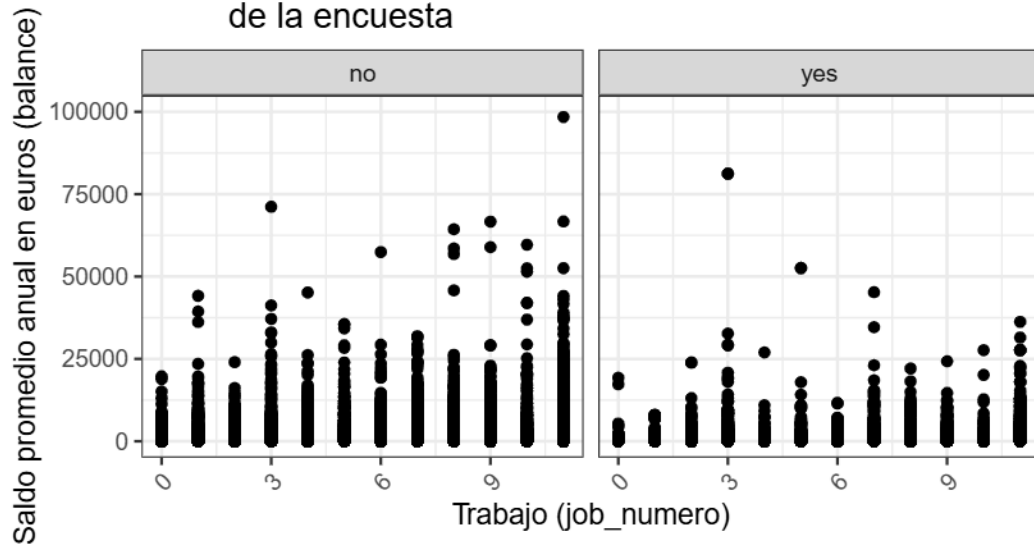
```

Warning: Use of `dataset_banco_clean\$job_numerico` is discouraged.
 i Use `job_numerico` instead.

Warning: Use of `dataset_banco_clean\$balance` is discouraged.
 i Use `balance` instead.

Warning: Removed 3767 rows containing missing values or values outside the scale range
 (`geom_point()`).

Resultado de la campaña de marketing anterior,
con respecto a las personas que se
suscribieron a una cuenta de deposito al termino
de la encuesta



Se analizará las variables “job_numerico” y “balance” en el grupo de personas que aceptaron la cuenta de depósito con el banco.

```
#Se guarda en el vector vector_y_yes, todas job_numerico de las personas que aceptaron la cu
vector_job_numerico_yes <- vector(mode="numeric", length=cont1)
x <- 1;
for (i in 1:45189) {
  if (dataset_banco_clean$y_numerico[i] == 1){
    vector_job_numerico_yes[x] <- dataset_banco_clean$job_numerico[i]
    x <- x + 1;
  }
}
print("Se creo vector vector_job_numerico_yes")
```

```
[1] "Se creo vector vector_job_numerico_yes"
```

```
#ver vector vector_job_numerico_yes
head(vector_job_numerico_yes)
```

```
[1] 8 8 6 8 11 11
```

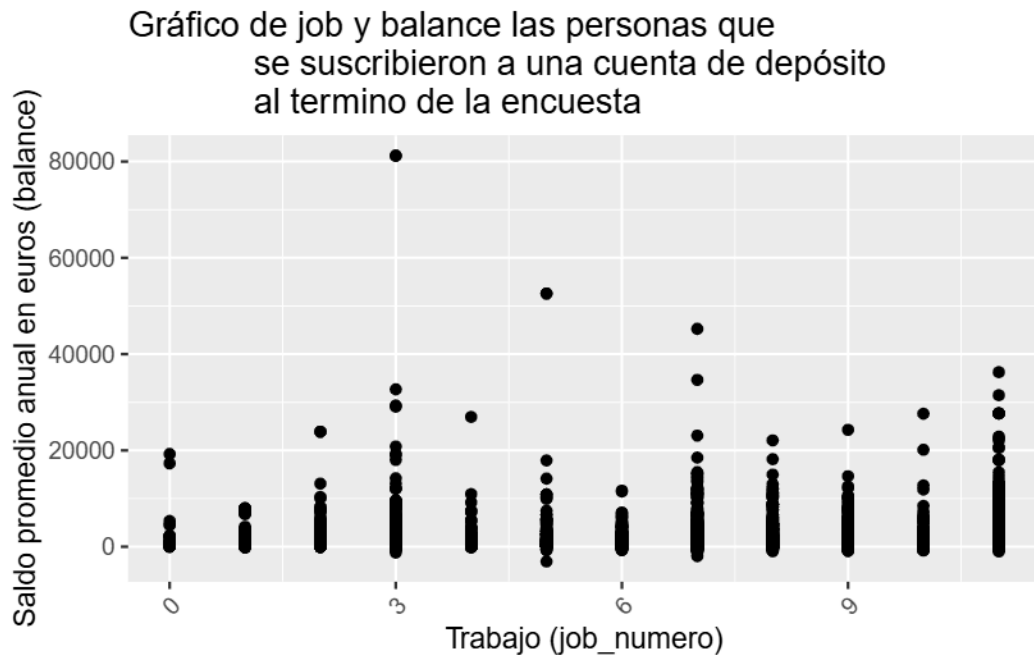
```
#Se guarda en el vector vector_balance_yes , todas balance de las personas que aceptaron la
vector_balance_yes <- vector(mode="numeric", length=cont1)
y <- 1;
for (j in 1:45189) {
  if (dataset_banco_clean$y_numerico[j] == 1){
    vector_balance_yes[y] <- dataset_banco_clean$balance[j]
    y <- y + 1;
  }
}
print("Se creo vector vector_balance_yes")
```

```
[1] "Se creo vector vector_balance_yes"
```

```
#ver vector vector_balance_yes
head(vector_balance_yes)
```

```
[1] 2343    45 2476  184    0  830
```

```
#Grafico de job y balance las personas que se suscribieron a una cuenta de deposito al termin
ggplot() +
  geom_point(mapping = aes(x = vector_job_numerico_yes, y = vector_balance_yes)) +
  theme(axis.text.x = element_text(angle = 45, hjust = 1)) +
  labs(
    title = "Gráfico de job y balance las personas que
            se suscribieron a una cuenta de depósito
            al termino de la encuesta",
    y = "Saldo promedio anual en euros (balance)",
    x = "Trabajo (job_numero)"
  )
```



Interpretación:

Se puede observar en el gráfico que las personas que se suscribieron a una cuenta de depósito, los empleos con mayor cantidad son los jubilados, administrativo y gerentes. Y son los que tienen un mayor saldo en euros en su cuenta.

Una interpretar porque los jubilados tiene mas ahorros y los otros 2 empleos poseen sueldo mas altos.

6. Conclusión

Primero para mi fue muy gratificante realizar este trabajo, por es un desafío muy emocionante debido a mis gusto personal por las matemáticas.

Realizando el trabajo me di cuenta de mi falta de conocimiento y practica.

Realice el análisis solo con 2 variables, job (trabajo) y balance (saldo en euros) de cada persona encuestadas. En donde los empleos con mayor cantidad son los jubilados, administrativo y gerentes. Y son los que tienen un mayor saldo en euros en su cuenta.

Una interpretar porque los jubilados tiene mas ahorros y los otros 2 empleos poseen sueldo mas altos.

Principalmente porque debo analizar las otras variables y no alcance a realizarlo.

El análisis realizado es solo comparando 2 variables.

Me gustaría realizar un análisis multivariable simultáneamente, desconozco si es posible realizarlo.

Me gusta mucho realizarlo, porque funciona como punto de partida, para conocer mis propias debilidades y fortalezas.

De este modo ir mejorando continuamente.