
# Universidad Técnica de Manabí


## Facultad de Ciencias Informáticas

## Carrera de Tecnología de la Información

**Integrantes:**


- *Ángel David Macías Ramírez*
- *Wilson Argandoña Macías*


**Materia:**

"Minería de Datos"

**Nivel**

6to

**Paralelo:**

"A"

*Ing. Ricardo Ordoñez*

***Resolución de problemas Unidad 1***
***Actividad 2***

**Actividad de Trabajo autónomo**

- **Material a realizar las actividades**

**DATASET_RP_1**
El dataset para la actividad de resolución de problemas 1, está enfocado en créditos de productos eléctricos y electrónicos de consumo para el hogar. Contiene 2987 registros para el análisis.

Para esta actividad tendremos la importación de las siguientes librerías.

```
#Cargamos las librerias
library("PASWR")
library("tidyverse")
library("readr")
library("ggplot2")
library("dplyr")
library("magrittr")
library("tibble")
library("dplyr")
library("caret")
library("devtools")
library("xlsx")
```

**Requerimientos de la actividad.**

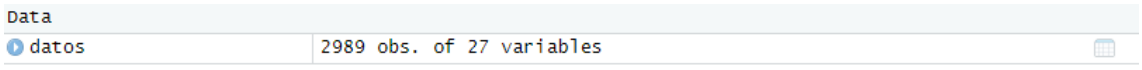
**Exploración de los datos:**

Antes de empezar, importaremos los datos del Excel, que hemos guardado en nuestra carpeta datos, con el siguiente comando

| **Código>** | datos <- read.xlsx("Datos/DATASET_RP_1.xlsx", sheetIndex = 1) |
|---|---|

```
#cargamos el dataset
datos <- read.xlsx("Datos/DATASET_RP_1.xlsx", sheetIndex = 1)
```

Resultado:




1. **Explorar las variables, ver el tipo de dato y ciertos datos.**
   Una vez, cargado el dataset, exploraremos los datos con el comando glimpse() integrando dentro el conjunto de datos.

| **Código>** | glimpse(datos) |
|---|---|

```
#1 Explorar las variables, ver el tipo de dato y ciertos datos
glimpse(datos)
```

Resultado:

```
> glimpse(datos)
Rows: 2,989
Columns: 27
$ Nombre.Empresa             <chr> "MINIMERCADO LUALJO", "MINIMERCADO LUALJO", "MINIMERCADO LUALJO", "MINIMERCADO LUALJO", "~
$ Periodo.cobro              <date> 2019-07-01, 2019-07-01, 2019-07-01, 2019-07-01, 2019-07-01, 2019-07-01, 2019-07-01, 2019~
$ Documento.Cliente          <chr> "39404956", "43916717", "43607729", "42964921", "98639592", "1017174588", "8061417", "719~
$ Nombre.Cliente             <chr> "Marleny Corrales Lopez", "Daniela Mejia Osorio", "Jhon Alexander Granada Zapata", "Brayh~
$ Telefono.Cliente           <chr> "3371032", NA, "2324836-23", "2284208", "4621848", "2645482", "2343927", "2143218", "2342~
$ Celular.Cliente            <dbl> 3162569497, NA, 3152188116, 3107569336, 3198415367, 3146238114, 3192645968, 3145693096, 3~
$ Producto                   <chr> "Computador", "Plancha de cabello", "Secadora", "Estufa", "Cafetera", "Estufa", "Portatil~
$ Marca                      <chr> "LG", "Haceb", "Samsung", "Whirpool", "Whirpool", "Haceb", "Samsung", "Electrolux", "ABA"~
$ Ciudad.Descripcion.Cliente <chr> "MEDELLIN", "MEDELLIN", "MEDELLIN", "MEDELLIN", "MEDELLIN", "MEDELLIN", "MEDELLIN", "MEDE~
$ Documento                  <chr> "126997", "129609", "133295", "133662", "135822", "136253", "136476", "138778", "138945",~
$ Fecha.deuda                <date> 2018-08-21, 2018-09-18, 2018-10-25, 2018-10-30, 2018-11-24, 2018-11-28, 2018-11-29, 2018~
$ Fecha.Vencimiento          <date> 2018-11-19, 2018-12-17, 2019-02-22, 2019-04-28, 2019-03-24, 2019-03-28, 2019-05-28, 2019~
$ Plazo                      <dbl> 90, 90, 120, 180, 120, 120, 180, 90, 60, 30, 150, 60, 150, 150, 60, 90, 120, 30, 120, 90,~
$ Total.credito              <dbl> 88260, 72222, 154238, 98388, 29111, 144455, 376667, 81400, 312125, 81111, 354000, 715000,~
$ Cuotas.canceladas          <dbl> 2, 1, 2, 6, 1, 0, 7, 5, 6, 1, 5, 8, 1, 0, 7, 8, 1, 4, 3, 2, 1, 0, 2, 5, 6, 8, 1, 5, 8, 4,~
$ Valor.Cartera              <dbl> 70608, 65000, 123390, 39355, 26200, 144455, 113000, 40700, 124850, 73000, 177000, 143000,~
$ Cuotas.pendientes          <dbl> 8, 9, 8, 4, 9, 10, 3, 5, 4, 9, 5, 2, 9, 10, 3, 2, 9, 6, 7, 8, 9, 10, 8, 5, 4, 2, 9, 5, 2,~
$ Dias.en.mora               <dbl> 224, 196, 129, 64, 99, 95, 34, 108, 137, 166, 44, 133, 43, 40, 128, 96, 66, 132, 36, 519,~
$ Rango.mora                 <chr> "Más de 180 días", "Más de 180 días", "De 121 a 150 días", "De 61 a 90 días", "De 91 a 12~
$ Documento.Vendedor         <chr> "1128394965", "71750790", "900449343", "71706004", "71660557", "1038626425", "98668026", ~
$ Nombre.Vendedor            <chr> "Juliana Andrea Palacio Durango", "Liliana Chaverra", "Esteban Cifuentes Diaz", "Jenny Ma~
$ Telefono.valido.de.contacto <chr> "Si", "No", "Si", "Si", "Si", "Si", "Si", "Si", "Si", "No", "Si", "Si", "Si", "Si", "Si",~
```

## 2. Ver las primeras 15 filas de la base de datos. Luego, mostrar las 5 últimas filas de la base de datos.

Ahora para explorar las primeras 15 filas, usaremos el comando head() que no permitirá mostrar las primeras filas, pero limitaremos esos datos a 15, asignando un (n=15) dentro de la función, quedando de la siguiente manera.

**Código>** head(datos, n=15)

```
#2 *****ver las primeras 15 filas*****
#mostramos con head el dataframe datos y lo limitamos a 15
head(datos, n=15)
```

Resultado:

```
> head(datos, n=15)
       Nombre.Empresa Periodo.cobro Documento.Cliente                Nombre.Cliente Telefono.Cliente Celular.Cliente
1  MINIMERCADO LUALJO    2019-07-01          39404956         Marleny Corrales Lopez          3371032      3162569497
2  MINIMERCADO LUALJO    2019-07-01          43916717           Daniela Mejia Osorio             <NA>              NA
3  MINIMERCADO LUALJO    2019-07-01          43607729 Jhon Alexander Granada Zapata       2324836-23      3152188116
4  MINIMERCADO LUALJO    2019-07-01          42964921                 Brayhan Sierra          2284208      3107569336
5  MINIMERCADO LUALJO    2019-07-01          98639592  Luis Alberto Barbutin Romero          4621848      3198415367
6  MINIMERCADO LUALJO    2019-07-01        1017174588                   Bruno paula          2645482      3146238114
7  MINIMERCADO LUALJO    2019-07-01           8061417                  Deyse Yadira          2343927      3192645968
8  MINIMERCADO LUALJO    2019-07-01          71993204                   Dines hati          2143218      3145693096
9  MINIMERCADO LUALJO    2019-07-01          39435615  Yeyby Yusbey Grosso Cardenas          2342949      3174735584
10 MINIMERCADO LUALJO    2019-07-01          21851723                 Nubia Pérez             <NA>              NA
11 MINIMERCADO LUALJO    2019-07-01        1017173956      Maria Jaqueline Giraldo          4723217      3154207434
12 MINIMERCADO LUALJO    2019-07-01          52863643       Katerine Marin Martinez          5856651      3122818000
13 MINIMERCADO LUALJO    2019-07-01          70106275   Javier Antonio Arroyo Gomez          4411263      3108293760
14 MINIMERCADO LUALJO    2019-07-01          43740972                 Robert Puert          5784894      3157536787
15 MINIMERCADO LUALJO    2019-07-01          43636370           Edgar Coneo Coneo           2678078      3151013356
             Producto      Marca Ciudad.Descripcion.Cliente Documento Fecha.deuda Fecha.Vencimiento Plazo Total.credito
1          Computador         LG                   MEDELLIN    126997  2018-08-21        2018-11-19     90         88260
2  Plancha de cabello      Haceb                   MEDELLIN    129609  2018-09-18        2018-12-17     90         72222
3            Secadora    Samsung                   MEDELLIN    133295  2018-10-25        2019-02-22    120        154238
```

Luego nos pide mostrar las ultimas 5 filas, para este caso usaremos la función **tail()**, que nos permite mostrar las últimas filas de nuestro conjunto de datos, y limitamos nuevamente, pero esta vez con valor a 5 (n=5)

**Código>** tail(datos, n=5)

```
# *****motramos la ultimas 5 filas*****
#comando tail muestra las ultimas filas, pero lo limitamos a 5, para mostrar las ultimas 5 filas

tail(datos, n=5)
```

Resultado:

```
> tail(datos, n=5)
      Nombre.Empresa Periodo.cobro Documento.Cliente                Nombre.Cliente Telefono.Cliente Celular.Cliente
2985 MINIMERCADO LUALJO    2019-07-01          21546503                Migdalia Gomez          4611254      3102864072
2986 MINIMERCADO LUALJO    2019-07-01          43926496 David Enrique Hernandez Durango          2672795      3138725739
2987 MINIMERCADO LUALJO    2019-07-01           9761568               Leopoldinaduarte          2695725      3103262782
2988 MINIMERCADO LUALJO    2019-07-01         430566815                         Mario          4245621      3136554257
2989 MINIMERCADO LUALJO    2019-07-01          42685954                Wilson Avelino          2588746      3100609795
              Producto      Marca Ciudad.Descripcion.Cliente Documento Fecha.deuda Fecha.Vencimiento Plazo Total.credito
2985 Aire acondicionado Electrolux                    BELLO    119420  2018-05-17        2018-08-15    90          2031
2986              Xbox  Microsoft                MEDELLIN    134540  2018-11-08        2019-01-07    60         95818
2987          Licuadora    Samsung                MEDELLIN    136240  2018-11-28        2019-05-27   180         14375
2988            secador         LG                MEDELLIN    126931  2018-08-27        2019-01-16   150         25541
2989          Secadora Electrolux                MEDELLIN    129047  2018-09-13        2018-10-13    30         97500
     Cuotas.canceladas Valor.Cartera Cuotas.pendientes Dias.en.mora         Rango.mora Documento.Vendedor
2985                 1          1828                 9          320    Más de 180 días         1128404709
2986                 6         38327                 4          175 De 151  a 180 días         1128394965
2987                 2         11500                 8           35    De 31 a 60 días           51750989
2988                12          2300                 2          200 De 151  a 180 días          900449571
2989                11         97500                10          261    Más de 180 días          900449343
                Nombre.Vendedor Telefono.valido.de.contacto Usuario.contesta     Código.de.finalización Valor.acuerdo
2985               Leonardo Tequia                       Si               Si No tiene capacidad de pago             0
2986 Juliana Andrea Palacio Durango                       Si               Si No tiene capacidad de pago             0
2987  Angie Paola Correa Artunduaga                       Si               Si No tiene capacidad de pago             0
2988         Esteban Cifuentes Diaz                       Si               No No tiene capacidad de pago           500
2989         Esteban Cifuentes Diaz                       Si               Si             Acuerdo de pago         96525
```

3. **Mostrar únicamente las columnas nombre cliente, fecha deuda, producto, total de crédito, plazo, cuotas canceladas, y rango mora.**
   En este apartado, usaremos una consulta con select, la cual se filtraran las columnas que nos están solicitando, por lo que quedaría de la siguiente manera:

**Código>** select(datos,Nombre.Cliente, Fecha.deuda, Producto, Total.credito, Plazo, Cuotas.canceladas, Rango.mora)

```
#3. *****Mostrar únicamente las columnas nombre cliente, fecha deuda, producto, total *****
# *****de crédito, plazo, cuotas canceladas, y rango mora. *****

#con select seleccionamos las columnas que deseamos mostrar con el orden establecido
select(datos,Nombre.Cliente, Fecha.deuda, Producto, Total.credito, Plazo, Cuotas.canceladas, Rango.mora)
```

Resultado:

```
> select(datos,Nombre.Cliente, Fecha.deuda, Producto, Total.credito, Plazo, Cuotas.canceladas, Rango.mora)
                      Nombre.Cliente Fecha.deuda           Producto Total.credito Plazo Cuotas.canceladas
1                Marleny Corrales Lopez  2018-08-21         Computador         88260    90                 2
2                  Daniela Mejia Osorio  2018-09-18 Plancha de cabello         72222    90                 1
3         Jhon Alexander Granada Zapata  2018-10-25           Secadora        154238   120                 2
4                      Brayhan Sierra     2018-10-30             Estufa         98388   180                 6
5          Luis Alberto Barbutin Romero  2018-11-24           Cafetera         29111   120                 1
6                         Bruno paula    2018-11-28             Estufa        144455   120                 0
7                       Deyse Yadira     2018-11-29           Portatil        376667   180                 7
8                         Dines hati     2018-12-15              Horno         81400    90                 5
9          Yeyby Yusbey Grosso Cardenas  2018-12-16   Equipo de sonido        312125    60                 6
10                        Nubia Pérez    2018-12-17 Plancha de cabello         81111    30                 1
11            Maria Jaqueline Giraldo    2018-12-19           Portatil        354000   150                 5
12             Katerine Marin Martinez   2018-12-20            Celular        715000    60                 8
13          Javier Antonio Arroyo Gomez  2018-12-20           PC Gamer        135000   150                 1
14                      Robert Puert     2018-12-23               Xbox        150000   150                 0
15                Edgar Coneo Coneo      2018-12-25           Cafetera        330167    60                 7
16                           Grace k     2018-12-27   Equipo de sonido        464750    90                 8
17             Juan Felipe Uribe Henao   2018-12-27           Cafetera        221789   120                 1
18           Ana Isabel Pineros Sanchez  2019-01-20           Secadora        139892    30                 4
19                     Jaime Garavito    2019-01-26       Play Station         29929   120                 3
20         Jesus Antonio Osorio Quiceno  2017-10-30       Play Station         53721    90                 2
21             Maria Dolores Gil Gomez   2017-10-30 Aire acondicionado        194054    60                 1
22                            David      2017-10-30             Nevera        101000   150                 0
23                        Angel Miro     2017-10-30   Equipo de sonido        187005    30                 2
24      Libia Maria Ramirez Saldarriaga  2017-10-30              Horno        193920    60                 5
```

4. **Mostrar todos los datos, de los créditos con cuotas canceladas mayores a 10.**
   En este paso, tuvimos un inconveniente, no encontramos valores mayores a 10 en cuotas canceladas, por lo que tuvimos que agregar manualmente, por lo tanto agregamos 2 valores que son los que se mostraran en los resultados, para este paso usamos la función de filtro, el cual nos busca la condición que estamos asignando, luego mostramos el filtro.

**Código>** filtro <- filter(datos,datos$Cuotas.canceladas > 10);  view(filtro)

```r
#4 *****Mostrar todos los datos, de los créditos con cuotas canceladas mayores a 10.*****
# con el comando filter, filtramos los datos con el dataframe datos, para llamarlos
filtro <- filter(datos,datos$Cuotas.canceladas > 10)

view(filtro)
```

Resultado:

| | |
|---|---|
| datos | 2989 obs. of 27 variables |
| filtro | 2 obs. of 27 variables |

Filter

| Marca | Ciudad.Descripcion.Cliente | Documento | Fecha.deuda | Fecha.Vencimiento | Plazo | Total.credito | Cuotas.canceladas | Valor.Cartera | Cuotas.pendientes | Dia |
|---|---|---|---|---|---|---|---|---|---|---|
| LG | MEDELLIN | 126931 | 2018-08-27 | 2019-01-16 | 150 | 25541 | 12 | 2300 | 2 | |
| Electrolux | MEDELLIN | 129047 | 2018-09-13 | 2018-10-13 | 30 | 97500 | 11 | 97500 | 10 | |

**Limpieza y transformación de los datos:**

5. **Ordenar la base de datos por total de crédito en orden descendente.**
   Para este paso, usaremos lo que es arrange(), una función que nos permite ordenar, ya sea de forma ascendente o descendente, como nos pide de forma descendente, usamos “desc”, y por ultimo view() para ver los resultados, quedando de la siguiente manera:

**Código>** dts_ordenados <- arrange(datos,desc(datos$Total.credito)) ; view(dts_ordenados)

```r
#5. *****Ordenar la base de datos por total de crédito en orden descendente.*****
  #con el comando arrange, seleccionamos la base de datos, y su variable total.credito
dts_ordenados <- arrange(datos,desc(datos$Total.credito))

#y para observar que han sido ordenados ponemos
view(dts_ordenados)
```

Resultado:

| | |
|---|---|
| datos | 2989 obs. of 27 variables |
| dts_ordenados | 2989 obs. of 27 variables |
| filtro | 2 obs. of 27 variables |

| | Marca | Ciudad.Descripcion.Cliente | Documento | Fecha.deuda | Fecha.Vencimiento | Plazo | Total.credito | Cuotas.canceladas | Valor.Cartera | Cuotas.pendientes |
|---|---|---|---|---|---|---|---|---|---|---|
| n | Sony | MEDELLIN | 000004 | 2017-10-30 | 2018-03-29 | 150 | 3118500 | 9 | 311850 | |
| | ABA | MEDELLIN | 000004 | 2017-10-30 | 2017-12-29 | 60 | 3005900 | 9 | 300590 | |
| | Sony | MEDELLIN | 000004 | 2017-10-30 | 2017-11-29 | 30 | 2947300 | 9 | 294730 | |
| | ABA | MEDELLIN | 000004 | 2017-10-30 | 2017-11-29 | 30 | 2496540 | 9 | 249654 | |
| | LG | MEDELLIN | 000004 | 2017-10-30 | 2018-02-27 | 120 | 2344840 | 9 | 234484 | |
| | LG | MEDELLIN | 000004 | 2017-10-30 | 2018-01-28 | 90 | 2238700 | 9 | 223870 | |
| | ABA | MEDELLIN | 000001 | 2017-10-30 | 2018-01-28 | 90 | 2206570 | 9 | 220857 | |
| casa | Whirpool | MEDELLIN | 000004 | 2017-10-30 | 2018-04-28 | 180 | 2207780 | 9 | 220778 | |
| | Electrolux | MEDELLIN | 000004 | 2017-10-30 | 2018-02-27 | 120 | 2186400 | 9 | 218640 | |
| | LG | MEDELLIN | 000004 | 2017-10-30 | 2017-12-29 | 60 | 2143300 | 9 | 214330 | |
| | Whirpool | MEDELLIN | 133739 | 2018-10-30 | 2018-12-29 | 60 | 2140665 | 6 | 856266 | |
| | Whirpool | MEDELLIN | 000004 | 2017-10-30 | 2018-02-27 | 120 | 2139610 | 9 | 213961 | |
| | ABA | MEDELLIN | 000004 | 2017-10-30 | 2018-02-27 | 120 | 2097450 | 9 | 209745 | |
| cabello | LG | MEDELLIN | 000004 | 2017-10-30 | 2018-04-28 | 180 | 2051300 | 9 | 205130 | |
| cabello | LG | MEDELLIN | 000004 | 2017-10-30 | 2017-11-29 | 30 | 2047900 | 9 | 204790 | |
| | Whirpool | MEDELLIN | 000001 | 2017-10-30 | 2018-02-27 | 120 | 2033200 | 9 | 203320 | |

Showing 1 to 16 of 2,989 entries, 27 total columns

**6. En la columna cuotas canceladas, son considerados valores inusuales aquellos valores superiores a 10 cuotas, debido a las políticas de crédito de la tienda. Por tanto, se desea crear una nueva columna reemplazando estos como valores faltantes. Reemplazar estos datos con el valor "NA".**

Para este apartado, nos solicita cambiar los valores de "Cuotas.Canceladas", cuyos valores son mayores de 10 por "NA", por ello aplicamos una mutación o modificación con la función Mutate() para crear un auxiliar llamado "crédito limpio", a este aplicamos la función replace(), la cual tomara como condición que sea mayor que 10, aplicando este cambio a la columna cuotas.canceladas.

**Código>** datos <- datos %>% mutate(creditoLimpio = Cuotas.canceladas); datos$Cuotas.canceladas <- replace(datos$creditoLimpio, datos$creditoLimpio>10, "NA")

```
#6. *****En la columna cuotas canceladas, son considerados valores inusuales aquellos valores superiores a 10 cuotas,
  #debido a las políticas de crédito de la tienda. Por tanto, se desea crear una nueva columna reemplazando estos como valore
  #faltantes. Reemplazar estos datos con el valor "NA".*****

  #creamos una columna llamada credito limpio usando la funcion mutate y le asignamos lo mismo que tiene Cuotas.canceladas
datos <- datos %>% mutate(creditoLimpio = Cuotas.canceladas)

  #modificamos la columna creditoLimpio creada pero le condicionamos > 6 se ponemos NA
datos$Cuotas.canceladas <- replace(datos$creditoLimpio, datos$creditoLimpio>10, "NA")
```

Resultado:

| Marca | Ciudad.Descripcion.Cliente | Documento | Fecha.deuda | Fecha.Vencimiento | Plazo | Total.credito | Cuotas.canceladas | Valor.Cartera | Cuotas.pendientes | D |
|---|---|---|---|---|---|---|---|---|---|---|
| Whirpool | MEDELLIN | 138119 | 2018-12-11 | 2019-03-11 | 90 | 6090 | 6 | 2436 | 4 | |
| Haceb | MEDELLIN | 096872 | 2017-11-23 | 2018-01-22 | 60 | 16686 | 3 | 11680 | 7 | |
| Electrolux | MEDELLIN | 099605 | 2017-12-14 | 2018-01-13 | 30 | 14375 | 2 | 11500 | 8 | |
| Whirpool | MEDELLIN | 110550 | 2018-02-05 | 2018-04-06 | 60 | 21437 | 7 | 6431 | 3 | |
| Sony | MEDELLIN | 116465 | 2018-04-02 | 2018-05-02 | 30 | 12778 | 1 | 11500 | 9 | |
| Electrolux | ENVIGADO | 116735 | 2018-04-05 | 2018-08-03 | 120 | 57500 | 8 | 11500 | 2 | |
| Haceb | MEDELLIN | 117377 | 2018-04-15 | 2018-06-14 | 60 | 12778 | 1 | 11500 | 9 | |
| Electrolux | MEDELLIN | 118204 | 2018-04-29 | 2018-07-28 | 90 | 23000 | 5 | 11500 | 5 | |
| Microsoft | MEDELLIN | 119028 | 2018-05-10 | 2018-08-08 | 90 | 83 | 4 | 50 | 6 | |
| Electrolux | BELLO | 119420 | 2018-05-17 | 2018-08-15 | 90 | 2031 | 1 | 1828 | 9 | |
| Microsoft | MEDELLIN | 134540 | 2018-11-08 | 2019-01-07 | 60 | 95818 | 6 | 38327 | 4 | |
| Samsung | MEDELLIN | 136240 | 2018-11-28 | 2019-05-27 | 180 | 14375 | 2 | 11500 | 8 | |
| LG | MEDELLIN | 126931 | 2018-08-27 | 2019-01-16 | 150 | 25541 | NA | 2300 | 2 | |
| Electrolux | MEDELLIN | 129047 | 2018-09-13 | 2018-10-13 | 30 | 97500 | NA | 97500 | 10 | |

**7. Crear una nueva columna, de nombre H_CREDITO, que contenga los valores: "habilitado para crédito" si las cuotas canceladas son mayores a 6, caso contrario el valor será "no habilitado para crédito".**

Para este paso, crearemos la nueva columna, seguido del resultado de la función que se pide, que es cuando las cuotas.canceladas sean mayores a 6 se se habilitara el crédito, caso contrario no se le habilitara el crédito. Para esta creación y modificación usaremos la función mutate(), y para la condición usaremos Case_when().

**Código>** datos <- datos %>% mutate(H_CREDITO = case_when(Cuotas.canceladas > 6 ~ "habilitado para el credito", negate = TRUE ~ "no habilitao para el credito"))

Para verificar que los datos son correctos, usamos una consulta select().

**Código>** select(datos, Cuotas.canceladas,H_CREDITO)

```r
#7 *****Crear una nueva columna, de nombre H_CREDITO, que contenga los valores: "habi
   #el valor será "no habilitado para crédito". *****

  #usamos mutate para crear una columna H_Credito con case_when como condicion cuotra
  #y por falso ponemos no habilitado para el credito
datos <- datos %>%
  mutate(H_CREDITO = case_when(Cuotas.canceladas > 6 ~ "habilitado para el credito",
                               negate = TRUE ~ "no habilitao para el credito"))
#verificar datos para saber que estan correctamente ordenados
select(datos, Cuotas.canceladas,H_CREDITO)
```

Resultado:

```
> select(datos, Cuotas.canceladas,H_CREDITO)
    Cuotas.canceladas                    H_CREDITO
1                   2 no habilitao para el credito
2                   1 no habilitao para el credito
3                   2 no habilitao para el credito
4                   6 no habilitao para el credito
5                   1 no habilitao para el credito
6                   0 no habilitao para el credito
7                   7   habilitado para el credito
8                   5 no habilitao para el credito
9                   6 no habilitao para el credito
10                  1 no habilitao para el credito
11                  5 no habilitao para el credito
12                  8   habilitado para el credito
13                  1 no habilitao para el credito
14                  0 no habilitao para el credito
15                  7   habilitado para el credito
16                  8   habilitado para el credito
17                  1 no habilitao para el credito
18                  4 no habilitao para el credito
19                  3 no habilitao para el credito
20                  2 no habilitao para el credito
21                  1 no habilitao para el credito
22                  0 no habilitao para el credito
23                  2 no habilitao para el credito
```

**8. Se requiere que el nombre de un producto en específico se encuentre correctamente escrito. Existen productos ingresados como Audifono, audifono y audifon, secado, secador, cecadora. Crear una nueva columna que contenga el dato adecuado establecido como "Audifonos" y "Secadora", para todos los registros**.

Para este paso, nos solicita una corrección en los nombres, incompletos o mal escrito, para esta función, usamos un mutate(), donde modificaremos los nombres erróneos, como Audifono, audífono, audifon por la palabra correcta "Audifonos", así mismo con la palabra secadora. Quedando de la siguiente manera:

**Código>** datos <- datos %>% mutate(

Producto = str_replace_all(Producto, "Audifono", "Audifonos"),

Producto = str_replace_all(Producto, "Audifonoss", "Audifonos"),

Producto = str_replace_all(Producto, "audifono", "Audifonos"),

Producto = str_replace_all(Producto, "audifon", "Audifonos"),

Producto = str_replace_all(Producto, "secador", "Secadora"),

```
  Producto = str_replace_all(Producto, "secado", "Secadora"),
  Producto = str_replace_all(Producto, "cecadora", "Secadora")
)
```

```
  #modificaremos con mutate la columna producto, para ello usamos str_replace_all y cambiamos la palabras Audifono -audifono
  #al crear Audifono(creara Audifonoss, para ello damos para corregmios poniendo una nueva condicion), lo mismo con la palabr
  #secadora, encontraremos secado, secador, por lo que cambiaremos esos valores, con Secadora.
datos <- datos %>% mutate(
  Producto = str_replace_all(Producto, "Audifono", "Audifonos"),
  Producto = str_replace_all(Producto, "Audifonoss", "Audifonos"),
  Producto = str_replace_all(Producto, "audifono", "Audifonos"),
  Producto = str_replace_all(Producto, "audifon", "Audifonos"),
  Producto = str_replace_all(Producto, "secador", "Secadora"),
  Producto = str_replace_all(Producto, "secado", "Secadora"),
  Producto = str_replace_all(Producto, "cecadora", "Secadora")
)
```

**Resultado:**

**Antes de aplicar el filtro.**

| Celular.Cliente | Producto | Marca | Ciudad.Descripcic |
|---|---|---|---|
| 3134561104 | audifon | Whirpool | MEDELLIN |
| 3153228161 | audifon | ABA | MEDELLIN |
| 3175874647 | audifon | Whirpool | MEDELLIN |
| 3119268719 | audifono | Haceb | MEDELLIN |
| 3168334754 | audifono | ABA | MEDELLIN |
| 3111776474 | audifono | LG | MEDELLIN |
| 3120426117 | audifono | LG | MEDELLIN |
| 3182314279 | audifono | Samsung | MEDELLIN |
| 3148027171 | audifono | ABA | MEDELLIN |
| 3107300559 | audifono | Samsung | MEDELLIN |
| 3188976162 | audifono | Electrolux | MEDELLIN |
| 3163183214 | audifono | Samsung | MEDELLIN |
| 3160836792 | audifono | Whirpool | MEDELLIN |
| 3183692058 | audifono | Electrolux | MEDELLIN |
| 3164462909 | audifono | Samsung | MEDELLIN |
| 3103476146 | audifono | Haceb | MEDELLIN |

| Celular.Cliente | Producto | Marca |
|---|---|---|
| 3171060803 | secado | Samsung |
| 3167851688 | secado | Electrolux |
| 3197049262 | secado | ABA |
| 3112659836 | secado | LG |
| 3188328218 | secado | ABA |
| 3146219339 | secador | LG |
| 3171284643 | secador | Samsung |
| 3185096567 | secador | Haceb |
| 3102726056 | secador | Haceb |
| 3101665784 | secador | LG |
| 3178116070 | secador | ABA |
| 3110361765 | secador | LG |
| 3153183970 | secador | Samsung |
| 3149190451 | secador | Haceb |
| 3152053364 | secador | Haceb |

Despues de aplicar el filtro.

| Celular.Cliente | Producto | Marca |
|---|---|---|
| 3135224358 | Aire acondicionado | LG |
| 3196211861 | Aire acondicionado | ABA |
| 3102864072 | Aire acondicionado | Electrolux |
| 3144048124 | Audifonos | Whirpool |
| 3165055031 | Audifonos | ABA |
| 3146897076 | Audifonos | ABA |
| 3180196046 | Audifonos | Electrolux |
| 3103476146 | Audifonos | Haceb |
| 3182498386 | Audifonos | Electrolux |
| 3176775096 | Audifonos | Samsung |
| 3104270469 | Audifonos | Haceb |
| 3106335771 | Audifonos | Haceb |
| 3184590111 | Audifonos | LG |
| 3189670113 | Audifonos | Electrolux |
| 3185285170 | Audifonos | ABA |
| 3140987769 | Audifonos | Whirpool |

| Celular.Cliente | Producto | Marca |
|---|---|---|
| *NA* | Portatil | ABA |
| 3148722293 | Portatil | ABA |
| 3109933267 | Portatil | Haceb |
| 3157668742 | Portatil | Electrolux |
| 3152188116 | Secadora | Samsung |
| 3199327332 | Secadora | ABA |
| 3116440374 | Secadora | Electrolux |
| 3130361134 | Secadora | Samsung |
| 3169278734 | Secadora | Electrolux |
| 3178598070 | Secadora | Samsung |
| 3198131977 | Secadora | Whirpool |
| 3166826546 | Secadora | Whirpool |
| 3177924541 | Secadora | ABA |
| 3119845872 | Secadora | Electrolux |
| 3184623666 | Secadora | ABA |
| 3168688166 | Secadora | Samsung |

**Visualización de los datos:**

## 9. Crear un gráfico de barras para visualizar distribución de rango mora.

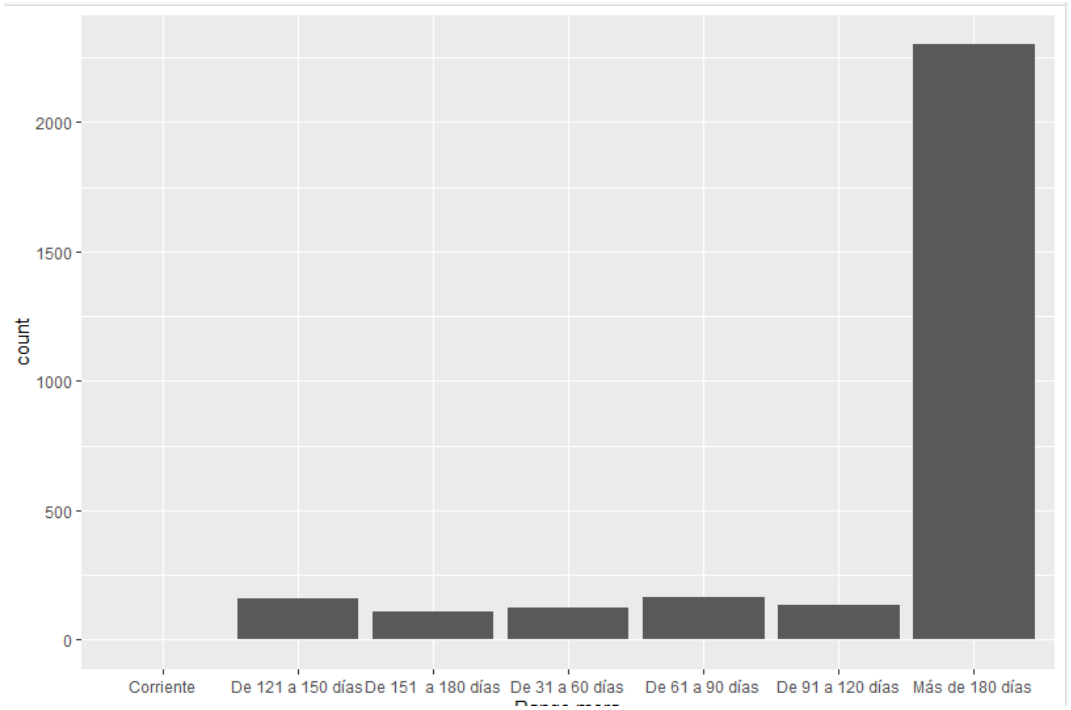
Aquí generaremos un gráfico, con ayuda de la función ggplot(), que nos permitirá graficar, dependiendo de la variable principal que deseemos graficar, en nuestro caso, usaremos geom_bar(), para generar un gráfico de barras y a la variable aplicada, “rango.mora” quedando de la siguiente manera.

**Código>** ggplot(data = datos) + geom_bar(aes(x = Rango.mora))

```
#9 Crear un gráfico de barras para visualizar distribución de rango mora.

  #usamos ggplot para crear los datos, como data(usamos el dataframe datos) y geom_bar para crear una grafica de barra
  # con rango Mora
ggplot(data = datos) + geom_bar(aes(x = Rango.mora))
```

**Resultado:**



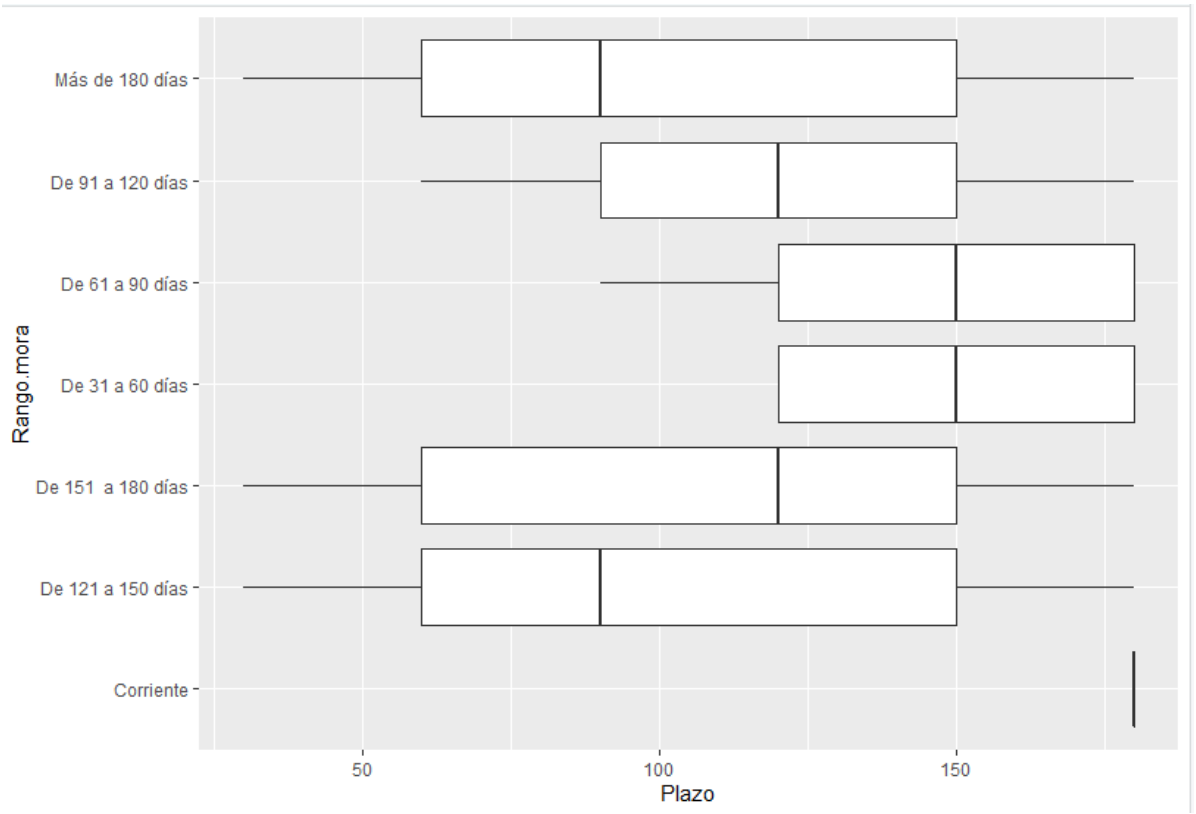
Como se muestra en el gráfico, existe rango de mora por “más de 180 dias”.

### 10. Dibujar un diagrama de cajas y bigotes para analizar la distribución de la variable plazo en función de la variable rango mora

Por ultimo nos pide otro gráfico, pero este será un diagrama llamado cajas y bigotes, este grafico utiliza dos variables para dedición que será: para X = Plazo y  Y = rango mora, y para que se ejecute el diagrama solicitado usamos geom_boxplot(), de la librería ggplot, dando como resultado:

**Código>** ggplot(data = datos) + geom_boxplot(aes(x = Plazo, y = Rango.mora))

```
#10. Dibujar un diagrama de cajas y bigotes para analizar la distribución de la variable
#plazo en función de la variable rango mora.
 #volvemos a usar ggplot para cargar el dataframe datos y usamos geom_boxplot para grafico de cajas y bigote,
 #con rango x de plazo y Y de mora
ggplot(data = datos) + geom_boxplot(aes(x = Plazo, y = Rango.mora))
```




## Conclusión.

- Estos puntos fueron muy útiles a la hora de elaborar un análisis, limpieza y trasformación de datos, por lo que los resultados deberían ser un poco más precisos de apreciar.
- Con una cantidad de datos mayor, se puede apreciar una limpieza más compleja pero más eficiente a la hora de obtener resultados.