Case6

Keyla Condor

2024-07-13

## Limpieza data

library(tidyverse)

## Warning: package 'tidyverse' was built under R version 4.1.3

## Warning: package 'tibble' was built under R version 4.1.3

## Warning: package 'forcats' was built under R version 4.1.3

## -- 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 3.5.0 v tibble 3.2.1  
## v lubridate 1.9.3 v tidyr 1.3.1  
## v purrr 1.0.2   
## -- 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 errors

library(rio)  
library(dplyr)  
library(readr)

Como primer paso debemos exportar la data

suscripcion<-read.csv("base\_suscripciones.csv")

# Cambio del n.a

Se reemplaza los n.a por “sin dato” como se muestra.

suscripcion <- mutate\_all(suscripcion,~replace(., is.na(.), "sin dato"))

# Filtro solo activas

ahora solo filtramos las suscripciones activas

suscripciones\_soloactivas <- suscripcion %>%  
 mutate(  
 fecha\_fin = ifelse(estado == "Activa", NA, fecha\_fin)  
 ) %>%  
 filter(is.na(fecha\_fin))

# Antiguedad

Luego se procede a crear una columna que contenga la variable antiguedad en años

suscripcion$fecha\_inicio <- as.Date(suscripcion$fecha\_inicio, format = "%Y-%m-%d")  
suscripcion$fecha\_fin <- as.Date(suscripcion$fecha\_fin, format = "%Y-%m-%d")  
  
suscripcion <- suscripcion %>%  
 mutate(  
 antiguedad = ifelse(is.na(fecha\_fin), NA, as.integer(difftime(fecha\_fin, fecha\_inicio, units = "days") / 30))  
 )

# suscripciones Activa na

hacemos una tabla donde se ponen a las suscripciones donde tienen la fecha na

suscripcion\_activas <- suscripcion %>%  
 mutate(fecha\_fin = ifelse(estado == "Activa", NA, fecha\_fin))

# Fechas correctas

Aqui procedemmos a separar por las que tienen el formato de la fecha y las que no poseen o son incorrectas.

is\_valid\_date <- function(date) {  
 grepl("^\\d{4}-\\d{2}-\\d{2}$", date)  
}  
  
fecha\_incorrectas <- suscripcion %>%  
 filter(!is\_valid\_date(fecha\_inicio) | !is\_valid\_date(fecha\_fin))  
  
suscripcion <- suscripcion %>%  
 filter(is\_valid\_date(fecha\_inicio) & is\_valid\_date(fecha\_fin))

# Separacion na

En este procedemos se juntan los datos na en otra tabla

registros\_no\_encontrados <- suscripcion %>%  
 filter(nombre\_cliente %in% c("", "sin dato") | fecha\_fin %in% c("", "sin dato")  
 | estado %in% c("","sin dato"))

# Solo positivos

Aqui separamos los negativos y positivos

valores\_negativos <- suscripcion %>%  
 filter(id\_suscripcion< 0)  
  
suscripcion<- suscripcion[suscripcion$id\_suscripcion>0,]

# Fechas correctas

Donde las fechas si coordinan y la fecha fin es mayor a la de inicio

suscripcion\_fecha\_mal<- suscripcion %>%  
 filter(fecha\_inicio >= fecha\_fin | is.na(fecha\_fin))  
  
suscripcion<- suscripcion %>%  
 filter(fecha\_fin >= fecha\_inicio | is.na(fecha\_fin))

# Valores incorrectos

agrupacion de todos los valores incorrectos

valores\_incorrectos <- bind\_rows(  
 valores\_negativos,  
 registros\_no\_encontrados,  
 fecha\_incorrectas,  
 suscripcion\_fecha\_mal  
)

# data limpia

print(suscripcion)

## id\_suscripcion id\_cliente nombre\_cliente fecha\_inicio fecha\_fin estado  
## 1 72 50 Ana 2022-04-14 2023-11-15 Activa  
## 2 44 9 Carlos 2018-03-27 2023-06-12 Cancelada  
## 3 23 30 sin dato 2018-08-11 2023-11-05 Activa  
## 4 89 43 Juan 2018-06-11 2020-02-13 Activa  
## 5 88 22 Ana 2018-10-04 2021-08-27 Expirada  
## 6 89 1 Luis 2018-03-31 2019-10-17 Activa  
## 7 59 37 sin dato 2018-11-30 2021-04-10 Expirada  
## 8 94 32 Juan 2018-03-27 2023-09-03 Activa  
## 9 54 16 MarÃ­a 2018-11-12 2023-08-17 Activa  
## 10 18 21 Luis 2020-08-15 2021-11-24 Cancelada  
## 11 95 17 Ana 2021-10-05 2023-02-10 Activa  
## 12 60 32 Carlos 2018-09-08 2022-12-16 Expirada  
## 13 16 50 Ana 2020-07-29 2021-04-29 Cancelada  
## 14 63 43 sin dato 2018-05-08 2021-10-13 Cancelada  
## 15 78 41 Carlos 2022-10-04 2022-12-05 Activa  
## 16 50 13 MarÃ­a 2021-01-10 2022-07-29 sin dato  
## 17 14 48 Ana 2020-01-02 2023-06-18 Activa  
## 18 46 15 sin dato 2018-06-26 2019-05-19 Cancelada  
## 19 87 32 Luis 2019-07-19 2021-05-16 Expirada  
## 20 38 26 Juan 2019-12-11 2023-04-26 Activa  
## 21 72 1 Luis 2018-01-08 2023-03-05 Activa  
## 22 56 2 Carlos 2018-07-10 2021-12-06 sin dato  
## 23 55 6 sin dato 2018-12-20 2021-05-23 sin dato  
## 24 16 44 sin dato 2019-07-03 2022-10-12 sin dato  
## 25 44 5 MarÃ­a 2021-04-27 2022-07-26 Expirada  
## 26 35 31 Ana 2018-01-11 2020-05-21 Expirada  
## 27 18 44 MarÃ­a 2019-06-03 2020-11-05 sin dato  
## 28 89 24 MarÃ­a 2020-05-05 2021-01-12 Activa  
## 29 3 48 sin dato 2018-07-16 2023-07-06 Activa  
## 30 23 14 Carlos 2018-05-12 2019-11-29 Activa  
## 31 58 30 Juan 2019-10-02 2023-08-30 sin dato  
## 32 66 24 MarÃ­a 2020-12-30 2023-10-23 Expirada  
## 33 85 12 Luis 2018-02-28 2020-05-09 Activa  
## 34 19 50 Ana 2018-12-21 2022-02-14 Activa  
## 35 28 33 Juan 2021-10-12 2022-01-13 sin dato  
## 36 7 30 Juan 2020-12-03 2023-09-25 sin dato  
## 37 5 27 Juan 2018-07-11 2023-08-12 sin dato  
## 38 38 13 Carlos 2018-06-08 2022-12-17 Cancelada  
## 39 23 35 sin dato 2020-10-19 2022-09-27 Activa  
## 40 92 18 Carlos 2021-06-15 2023-05-09 Cancelada  
## 41 40 35 Ana 2022-11-10 2022-11-12 Invalido  
## 42 86 23 Carlos 2021-04-01 2021-08-23 Expirada  
## 43 16 40 Ana 2021-03-01 2023-01-24 Expirada  
## 44 84 36 MarÃ­a 2021-04-04 2023-10-12 Cancelada  
## 45 27 26 Luis 2019-01-08 2019-11-26 Cancelada  
## 46 15 32 Carlos 2018-04-30 2021-08-14 Cancelada  
## 47 17 40 Carlos 2020-06-07 2022-10-25 Cancelada  
## 48 8 4 Ana 2018-12-11 2021-07-30 Activa  
## 49 52 29 Ana 2018-01-10 2020-08-28 Activa  
## 50 37 25 Carlos 2022-08-07 2023-08-08 Activa  
## 51 7 23 MarÃ­a 2021-10-10 2022-12-12 Activa  
## 52 53 8 Luis 2019-07-01 2019-12-19 Expirada  
## 53 25 36 sin dato 2020-05-19 2021-02-10 Cancelada  
## 54 59 28 Juan 2020-02-24 2023-06-25 Expirada  
## 55 51 3 Juan 2022-05-23 2022-11-19 Activa  
## 56 51 47 Luis 2018-11-06 2021-03-03 sin dato  
## 57 8 12 MarÃ­a 2018-03-26 2019-03-03 Expirada  
## 58 78 3 sin dato 2020-06-14 2021-02-12 Expirada  
## 59 37 45 Carlos 2020-04-10 2022-10-28 Expirada  
## 60 77 9 Juan 2018-05-01 2021-10-07 sin dato  
## 61 22 44 Ana 2019-10-18 2021-11-19 sin dato  
## 62 37 30 Ana 2020-02-28 2021-11-14 Expirada  
## 63 59 14 MarÃ­a 2018-09-25 2020-01-15 Cancelada  
## 64 12 16 Carlos 2021-03-03 2021-10-28 Expirada  
## 65 75 20 MarÃ­a 2020-02-17 2020-03-28 sin dato  
## 66 33 9 Ana 2018-06-24 2019-01-22 sin dato  
## 67 32 33 Juan 2022-03-16 2022-12-19 Activa  
## 68 22 30 Carlos 2020-07-29 2021-05-31 Cancelada  
## 69 83 42 MarÃ­a 2022-04-20 2022-05-23 sin dato  
## 70 38 23 Juan 2019-11-24 2023-09-05 Activa  
## 71 72 36 Juan 2019-03-07 2022-11-29 Cancelada  
## 72 23 42 Luis 2019-02-24 2022-08-23 Invalido  
## 73 20 27 Ana 2019-11-23 2023-03-02 Activa  
## 74 38 9 sin dato 2019-01-31 2020-03-26 Expirada  
## 75 4 9 Carlos 2018-05-18 2022-04-23 sin dato  
## 76 48 17 MarÃ­a 2021-09-13 2023-08-03 Activa  
## 77 80 46 sin dato 2018-12-25 2022-11-18 Cancelada  
## 78 88 36 Carlos 2018-10-23 2019-03-20 Expirada  
## 79 8 1 MarÃ­a 2021-01-21 2022-11-03 sin dato  
## 80 92 4 Juan 2020-12-22 2022-06-25 sin dato  
## 81 6 8 MarÃ­a 2022-12-27 2023-01-18 Activa  
## 82 10 47 Carlos 2018-12-31 2022-10-31 Cancelada  
## 83 10 16 sin dato 2021-08-04 2022-09-19 Expirada  
## 84 6 19 sin dato 2018-01-07 2019-07-05 sin dato  
## antiguedad  
## 1 19  
## 2 63  
## 3 63  
## 4 20  
## 5 35  
## 6 18  
## 7 28  
## 8 66  
## 9 57  
## 10 15  
## 11 16  
## 12 52  
## 13 9  
## 14 41  
## 15 2  
## 16 18  
## 17 42  
## 18 10  
## 19 22  
## 20 41  
## 21 62  
## 22 41  
## 23 29  
## 24 39  
## 25 15  
## 26 28  
## 27 17  
## 28 8  
## 29 60  
## 30 18  
## 31 47  
## 32 34  
## 33 26  
## 34 38  
## 35 3  
## 36 34  
## 37 61  
## 38 55  
## 39 23  
## 40 23  
## 41 0  
## 42 4  
## 43 23  
## 44 30  
## 45 10  
## 46 40  
## 47 29  
## 48 32  
## 49 32  
## 50 12  
## 51 14  
## 52 5  
## 53 8  
## 54 40  
## 55 6  
## 56 28  
## 57 11  
## 58 8  
## 59 31  
## 60 41  
## 61 25  
## 62 20  
## 63 15  
## 64 7  
## 65 1  
## 66 7  
## 67 9  
## 68 10  
## 69 1  
## 70 46  
## 71 45  
## 72 42  
## 73 39  
## 74 14  
## 75 47  
## 76 22  
## 77 47  
## 78 4  
## 79 21  
## 80 18  
## 81 0  
## 82 46  
## 83 13  
## 84 18