CASO4

Keyla condor

2024-07-12

## Limpieza de datos

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

polizas<-read.csv("base\_polizas.csv")

# Cambio del n.a

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

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

# Filtro solo activas

ahora solo filtramos las polizas activas

polizas\_solactivas <- polizas %>%  
 filter(!is.na(fecha\_fin) &   
 fecha\_fin >= fecha\_inicio &   
 estado == "Activa")

# Antiguedad

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

polizas <- polizas %>%  
 mutate(  
 fecha\_inicio = as.Date(fecha\_inicio, format = "%Y-%m-%d"),  
 fecha\_fin = as.Date(fecha\_fin, format = "%Y-%m-%d"),  
 antiguedad = ifelse(is.na(fecha\_fin), NA, as.integer(difftime(fecha\_fin, fecha\_inicio, units = "days") / 365))  
 )

# Polizas Activa na

recordar que tambien nos dice que el formato de la fecha tambien puede ser donde las polizas activas tienen na entonces creamos una tabla donde las polizas activas tienen el na y las demas respetan el formato.

polizas\_activas<- polizas %>%  
 mutate(fecha\_inicio = as.Date(fecha\_inicio, format = "%Y-%m-%d"),  
 fecha\_fin = as.Date(fecha\_fin, format = "%Y-%m-%d"))  
  
polizas\_activas <- polizas %>%  
 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 <- polizas %>%  
 filter(!is\_valid\_date(fecha\_inicio) | !is\_valid\_date(fecha\_fin))  
  
polizas <- polizas %>%  
 filter(is\_valid\_date(fecha\_inicio) & is\_valid\_date(fecha\_fin))

# Separacion na

En este procedemos a juntos los datos na en otra tabla

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

# Solo positivos

Aqui separamos los negativos y positivos

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

# Fechas correctas

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

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

# data limpia

print(polizas)

## id\_poliza nombre\_cliente fecha\_inicio fecha\_fin monto\_cobertura estado  
## 1 73 Juan 2015-09-16 2023-10-06 32523.54 Expirada  
## 2 95 Carlos 2018-04-13 2019-12-10 48975.72 Activa  
## 3 99 sin dato 2017-06-21 2021-06-05 11640.05 Cancelada  
## 4 3 Carlos 2017-02-15 2017-03-26 11716.5 sin dato  
## 5 32 Luis 2015-05-10 2022-08-29 12882.25 Expirada  
## 6 9 Juan 2017-04-19 2019-07-14 38716.08 Activa  
## 7 96 Juan 2018-12-25 2021-03-14 sin dato Activa  
## 8 99 Luis 2017-10-26 2023-04-24 82543.33 sin dato  
## 9 98 Ana 2015-01-09 2021-04-23 11782.84 Expirada  
## 10 53 sin dato 2017-09-16 2023-10-09 52519.73 Expirada  
## 11 49 Luis 2021-10-08 2023-03-11 76763.98 Cancelada  
## 12 14 sin dato 2018-11-29 2021-09-29 94875.41 Invalido  
## 13 76 Carlos 2018-01-31 2021-10-05 13243.73 Activa  
## 14 97 Luis 2017-10-13 2023-12-06 30134.1 sin dato  
## 15 45 Juan 2019-11-28 2021-07-25 66039.49 Cancelada  
## 16 31 Luis 2015-07-18 2018-03-15 36559.04 Expirada  
## 17 42 Carlos 2016-10-27 2017-03-23 29176.95 Expirada  
## 18 64 sin dato 2016-04-07 2017-04-09 33363.35 Expirada  
## 19 61 Luis 2019-04-27 2022-12-09 53196.91 sin dato  
## 20 61 Ana 2022-01-26 2023-04-03 96636.3 Activa  
## 21 58 sin dato 2018-12-15 2020-05-24 64791.57 Expirada  
## 22 85 Luis 2020-02-12 2020-11-21 52702.32 sin dato  
## 23 14 Carlos 2016-05-24 2019-12-08 76045.98 Expirada  
## 24 8 Ana 2017-08-09 2021-06-18 86593.84 Activa  
## 25 74 Luis 2017-06-17 2022-09-20 25374.43 sin dato  
## 26 37 Luis 2019-08-13 2022-07-01 93021.65 sin dato  
## 27 81 Juan 2020-03-28 2023-11-07 95191.96 Activa  
## 28 99 Luis 2016-01-01 2021-11-24 51309.75 Expirada  
## 29 26 sin dato 2015-03-30 2016-10-28 53684.5 sin dato  
## 30 74 Ana 2016-05-25 2018-08-13 37207.46 sin dato  
## 31 66 Ana 2019-07-05 2019-11-15 26225.46 Cancelada  
## 32 65 Carlos 2019-01-23 2021-02-23 31990.21 Expirada  
## 33 80 Luis 2021-09-23 2022-06-20 23091.65 sin dato  
## 34 93 Carlos 2017-03-12 2019-05-01 80362.64 Cancelada  
## 35 44 Carlos 2015-11-05 2023-06-16 24440.19 Expirada  
## 36 34 Ana 2016-09-12 2019-09-09 93258.26 Expirada  
## 37 73 MarÃ­a 2021-02-06 2023-04-12 31034.72 Cancelada  
## 38 37 sin dato 2018-11-28 2023-12-18 58464.82 Expirada  
## 39 2 Carlos 2016-05-03 2023-10-17 71076.37 Activa  
## 40 15 Luis 2022-02-19 2022-07-10 29679.9 sin dato  
## 41 9 MarÃ­a 2017-06-27 2018-09-20 44514.94 Cancelada  
## 42 7 Ana 2019-09-13 2020-07-01 42617.69 sin dato  
## 43 52 Ana 2020-11-04 2021-05-26 68833.55 sin dato  
## 44 56 Luis 2019-07-04 2019-10-14 46923.86 Activa  
## 45 26 sin dato 2015-09-20 2018-12-26 82535.71 Activa  
## 46 68 sin dato 2017-07-18 2023-09-06 95109.97 Activa  
## 47 83 Luis 2015-08-03 2018-12-02 85563.63 Expirada  
## 48 64 sin dato 2018-12-30 2020-07-12 39507.22 Cancelada  
## 49 92 Juan 2015-06-20 2020-03-28 47952.63 sin dato  
## 50 58 Luis 2015-07-12 2021-12-06 sin dato Expirada  
## 51 62 sin dato 2020-08-22 2023-02-08 64884.69 Activa  
## 52 25 Ana 2020-11-10 2023-04-03 63765.95 Activa  
## 53 85 Ana 2018-03-09 2020-03-14 30086.37 Expirada  
## 54 94 MarÃ­a 2016-07-21 2018-08-06 97181.02 Expirada  
## 55 31 Ana 2016-02-25 2018-07-06 13777.35 sin dato  
## 56 38 Carlos 2015-03-27 2017-07-15 63038.55 Activa  
## 57 89 Juan 2015-01-15 2016-10-23 50799.09 sin dato  
## 58 83 Ana 2017-03-05 2022-10-25 66498.45 sin dato  
## 59 87 Juan 2019-11-16 2021-01-29 69908.86 Cancelada  
## 60 80 Juan 2019-03-21 2021-01-22 12578.98 Activa  
## 61 68 Carlos 2016-06-07 2020-11-02 66940.7 Activa  
## 62 76 MarÃ­a 2018-12-17 2021-06-20 38732.24 Invalido  
## 63 12 MarÃ­a 2018-03-10 2020-09-24 21687.65 Expirada  
## 64 69 Luis 2016-07-25 2016-08-05 40952.43 Activa  
## 65 20 sin dato 2015-11-08 2021-06-12 55185.04 Cancelada  
## 66 68 Carlos 2022-06-21 2023-11-30 45266.96 sin dato  
## 67 59 Juan 2019-02-09 2019-07-20 sin dato Expirada  
## 68 57 Juan 2016-06-15 2023-01-05 21015.14 sin dato  
## 69 92 Juan 2017-07-15 2023-06-24 16663.42 Cancelada  
## 70 41 Juan 2017-08-08 2022-07-09 30254.93 Activa  
## 71 7 Luis 2015-06-04 2016-10-24 12432.04 Activa  
## 72 15 Carlos 2015-03-17 2021-08-13 17418.65 Cancelada  
## 73 82 Ana 2019-05-11 2019-07-28 11917.6 Activa  
## 74 38 Juan 2021-03-16 2022-09-11 18747.67 Cancelada  
## 75 1 Ana 2019-08-07 2023-08-04 72396.68 Cancelada  
## 76 14 Carlos 2018-05-26 2022-04-26 71425.56 Expirada  
## 77 52 Juan 2017-04-11 2022-08-17 80181.51 Activa  
## antiguedad  
## 1 8  
## 2 1  
## 3 3  
## 4 0  
## 5 7  
## 6 2  
## 7 2  
## 8 5  
## 9 6  
## 10 6  
## 11 1  
## 12 2  
## 13 3  
## 14 6  
## 15 1  
## 16 2  
## 17 0  
## 18 1  
## 19 3  
## 20 1  
## 21 1  
## 22 0  
## 23 3  
## 24 3  
## 25 5  
## 26 2  
## 27 3  
## 28 5  
## 29 1  
## 30 2  
## 31 0  
## 32 2  
## 33 0  
## 34 2  
## 35 7  
## 36 2  
## 37 2  
## 38 5  
## 39 7  
## 40 0  
## 41 1  
## 42 0  
## 43 0  
## 44 0  
## 45 3  
## 46 6  
## 47 3  
## 48 1  
## 49 4  
## 50 6  
## 51 2  
## 52 2  
## 53 2  
## 54 2  
## 55 2  
## 56 2  
## 57 1  
## 58 5  
## 59 1  
## 60 1  
## 61 4  
## 62 2  
## 63 2  
## 64 0  
## 65 5  
## 66 1  
## 67 0  
## 68 6  
## 69 5  
## 70 4  
## 71 1  
## 72 6  
## 73 0  
## 74 1  
## 75 3  
## 76 3  
## 77 5