

# *Caso4 - Pólizas*

Sebastian Matias Romero Davila

2024-07-10

## Contents

<i>Base de datos inicial</i>	<b>2</b>
<i>Características Inherentes</i>	<b>2</b>
Exactitud . . . . .	2
Compleitud . . . . .	3
Consistencia . . . . .	4
<i>Características Dependientes del sistema</i>	<b>6</b>
Disponibilidad . . . . .	6
Portabilidad . . . . .	6

## Base de datos inicial

Cargamos librerías a emplear

```
library(tidyverse)
library(rio)
library(kableExtra)
```

Base: base\_polizas

```
base_polizas <- import("base_polizas.csv")
head(base_polizas,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
73	Juan	2015-09-16	2023-10-06	32523.54	Expirada
57	Luis	2017-03-08	2016-01-24	20143.19	Expirada
46	Carlos	2021-10-29	2016-06-24	12556.60	Activa
95	Carlos	2018-04-13	2019-12-10	48975.72	Activa
81	Luis	2019-05-13	2018-04-03	31276.24	Activa
58	Luis	2017-05-25	2016-07-02	50242.70	NA
95	Ana	2018-09-24	2017-08-31	41559.31	Activa
61	María	2022-01-02	2021-10-24	38225.97	NA
60	Carlos	2019-03-20	2017-03-17	43685.26	NA
-59	Luis	2019-11-13	2020-11-01	39435.03	NA

## Características Inherentes

### Exactitud

Para la representación correcta de la realidad de los datos, empezamos identificando los tipos de variables

```
base_polizas %>% glimpse()
```

```
## Rows: 150
## Columns: 6
## $ id_poliza      <int> 73, 57, 46, 95, 81, 58, 95, 61, 60, -59, 99, 3, 32, 9, ~
## $ nombre_cliente <chr> "Juan", "Luis", "Carlos", "Carlos", "Luis", "Luis", "A~
## $ fecha_inicio   <chr> "2015-09-16", "2017-03-08", "2021-10-29", "2018-04-13"~
## $ fecha_fin      <IDate> 2023-10-06, 2016-01-24, 2016-06-24, 2019-12-10, 2018~
## $ monto_cobertura <dbl> 32523.54, 20143.19, 12556.60, 48975.72, 31276.24, 5024~
## $ estado         <chr> "Expirada", "Expirada", "Activa", "Activa", "Activa", ~
```

Luego, verificamos si existen inexactitudes

- id\_poliza: Se encuentran valores enteros negativos
- fecha\_inicio: Formato inadecuado

```

inexactospol <- base_polizas %>% filter(
  id_poliza<0|substr(fecha_inicio,3,3) == "-"
)
head(inexactospol,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
-59	Luis	2019-11-13	2020-11-01	39435.03	NA
-60	Ana	31-12-2023	2023-06-03	97722.68	NA
-70	Carlos	2016-10-17	2023-07-06	NA	Expirada
-84	NA	31-12-2023	2019-06-24	91929.45	Cancelada
-89	Luis	2020-12-26	2022-06-21	64217.63	Invalido
-9	María	31-12-2023	2017-04-18	NA	Cancelada
-28	Luis	2022-04-13	2022-03-03	28793.04	Cancelada
-30	Juan	31-12-2023	2023-10-02	92666.88	Cancelada
-75	Carlos	2022-11-09	2021-04-24	NA	Expirada
-4	NA	31-12-2023	2022-07-08	63631.95	Invalido

## Completitud

En la tabla se observan valores faltantes en varias de las filas.

```

polizas_faltantes <- base_polizas %>%
  filter(
    if_any(everything(), is.na)
  )
head(polizas_faltantes,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
58	Luis	2017-05-25	2016-07-02	50242.70	NA
61	María	2022-01-02	2021-10-24	38225.97	NA
60	Carlos	2019-03-20	2017-03-17	43685.26	NA
-59	Luis	2019-11-13	2020-11-01	39435.03	NA
99	NA	2017-06-21	2021-06-05	11640.05	Cancelada
3	Carlos	2017-02-15	2017-03-26	11716.50	NA
96	Juan	2018-12-25	2021-03-14	NA	Activa
99	Luis	2017-10-26	2023-04-24	82543.33	NA
-60	Ana	31-12-2023	2023-06-03	97722.68	NA
53	NA	2017-09-16	2023-10-09	52519.73	Expirada

Cantidad total de datos faltantes por cada columna en el conjunto de datos.

```

na_polizas <- colSums(is.na(base_polizas))
na_polizas

```

```
##      id_poliza  nombre_cliente  fecha_inicio  fecha_fin monto_cobertura
##              0              26              0              0              10
##      estado
##              39
```

Entonces, ajustaremos los valores faltantes (NA) de la siguiente manera:

nombre\_cliente: Se reemplazará con “Desconocido”.

estado: Se reemplazará con “Indeterminado”.

monto\_cobertura: Se asignará como -1 en caso de estar ausente.

```
base1 <- base_polizas %>% mutate(
  nombre_cliente = case_when(
    !is.na(nombre_cliente) ~ nombre_cliente,
    is.na(nombre_cliente) ~ "Desconocido"
  ),
  monto_cobertura = case_when(
    !is.na(monto_cobertura) ~ monto_cobertura,
    is.na(monto_cobertura) ~ -1
  ),
  estado = case_when(
    !is.na(estado) ~ estado,
    is.na(estado) ~ "Indeterminado"
  )
)
head(base1,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
73	Juan	2015-09-16	2023-10-06	32523.54	Expirada
57	Luis	2017-03-08	2016-01-24	20143.19	Expirada
46	Carlos	2021-10-29	2016-06-24	12556.60	Activa
95	Carlos	2018-04-13	2019-12-10	48975.72	Activa
81	Luis	2019-05-13	2018-04-03	31276.24	Activa
58	Luis	2017-05-25	2016-07-02	50242.70	Indeterminado
95	Ana	2018-09-24	2017-08-31	41559.31	Activa
61	María	2022-01-02	2021-10-24	38225.97	Indeterminado
60	Carlos	2019-03-20	2017-03-17	43685.26	Indeterminado
-59	Luis	2019-11-13	2020-11-01	39435.03	Indeterminado

## Consistencia

Previamente, se identificaron inconsistencias en las variables id\_poliza y nombre\_cliente, así como un problema con el formato de las variables nombre\_cliente, fecha\_inicio y fecha\_fin. Para abordar estas cuestiones, se realizaron las siguientes acciones:

```
base2 <- base1 %>% mutate(
  id_poliza = abs(id_poliza),
  nombre_cliente = as.character(nombre_cliente),
```

```

fecha_inicio = as.Date(case_when(
  substr(fecha_inicio,3,3) == "-" ~ paste( substr(fecha_inicio,7,10),
                                             substr(fecha_inicio,4,5),
                                             substr(fecha_inicio,1,2),
                                             sep = "-"),
  substr(fecha_inicio,3,3) != "-" ~ fecha_inicio)
),
fecha_fin = as.Date(fecha_fin),
monto_cobertura = round(monto_cobertura,2)) %>%
arrange(id_poliza)

head(base2,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
1	Ana	2019-08-07	2023-08-04	72396.68	Cancelada
2	Carlos	2016-05-03	2023-10-17	71076.37	Activa
3	Carlos	2017-02-15	2017-03-26	11716.50	Indeterminado
3	Juan	2021-03-08	2020-12-15	51035.84	Expirada
4	Desconocido	2023-12-31	2022-07-08	63631.95	Invalido
6	Desconocido	2022-01-26	2021-08-28	-1.00	Expirada
7	Ana	2019-09-13	2020-07-01	42617.69	Indeterminado
7	Luis	2015-06-04	2016-10-24	12432.04	Activa
8	Ana	2017-08-09	2021-06-18	86593.84	Activa
8	Carlos	2022-10-03	2017-04-23	79831.42	Activa

Finalmente, nos piden crear una variable para la antigüedad de la póliza en años.

```

base3 <- base2 %>% filter(monto_cobertura != -1) %>% mutate(
  fecha_temporal = fecha_inicio,
  fecha_inicio = case_when(
    fecha_inicio > fecha_fin ~ fecha_fin,
    TRUE ~ fecha_inicio
  ),
  fecha_fin = case_when(
    fecha_temporal > fecha_fin ~ fecha_temporal,
    TRUE ~ fecha_fin
  )
) %>% select(-fecha_temporal)
head(base3,10) %>%
kable(booktabs = TRUE,format = "latex") %>%
kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado
1	Ana	2019-08-07	2023-08-04	72396.68	Cancelada
2	Carlos	2016-05-03	2023-10-17	71076.37	Activa
3	Carlos	2017-02-15	2017-03-26	11716.50	Indeterminado
3	Juan	2020-12-15	2021-03-08	51035.84	Expirada
4	Desconocido	2022-07-08	2023-12-31	63631.95	Invalido
7	Ana	2019-09-13	2020-07-01	42617.69	Indeterminado
7	Luis	2015-06-04	2016-10-24	12432.04	Activa
8	Ana	2017-08-09	2021-06-18	86593.84	Activa
8	Carlos	2017-04-23	2022-10-03	79831.42	Activa
9	Juan	2017-04-19	2019-07-14	38716.08	Activa

También, filtrar las pólizas activas.

```
polizas_activas <- base3 %>% mutate(
  antiguedad_años = as.integer(trunc(difftime(fecha_fin, fecha_inicio)/360))
) %>% filter(estado == "Activa")
head(polizas_activas, 10) %>%
kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_poliza	nombre_cliente	fecha_inicio	fecha_fin	monto_cobertura	estado	antiguedad_años
2	Carlos	2016-05-03	2023-10-17	71076.37	Activa	7
7	Luis	2015-06-04	2016-10-24	12432.04	Activa	1
8	Ana	2017-08-09	2021-06-18	86593.84	Activa	3
8	Carlos	2017-04-23	2022-10-03	79831.42	Activa	5
9	Juan	2017-04-19	2019-07-14	38716.08	Activa	2
10	Ana	2017-06-10	2018-09-13	40754.09	Activa	1
17	Ana	2017-05-12	2022-02-15	64047.35	Activa	4
25	Ana	2020-11-10	2023-04-03	63765.95	Activa	2
26	Desconocido	2015-09-20	2018-12-26	82535.71	Activa	3
28	Ana	2016-11-30	2018-03-16	67659.53	Activa	1

## *Características Dependientes del sistema*

### Disponibilidad

Medida en que los datos están disponibles y accesibles cuando se necesitan.

[https://github.com/bastiannseef/E2422/tree/main/Caso4/data\\_procesada\\_caso4](https://github.com/bastiannseef/E2422/tree/main/Caso4/data_procesada_caso4)

### Portabilidad

Para trasladar los datos evaluados bajo el criterio de calidad de la norma ISO/IEC 25012 entre sistemas, guardamos las bases de datos procesadas en archivos “.csv” y también, en formato “.rds” para garantizar eficiencia, portabilidad y seguridad.

Encontrará los siguientes archivos:

- data\_caso4.csv: Data general evaluada por los criterios de calidad en formato “.csv”.
- data\_caso4.rds: Data general evaluada por los criterios de calidad en formato “.rds”.
- polizas\_activas.csv: Data pólizas activas en formato “.csv”.
- polizas\_activas.rds: Data pólizas activas en formato “.rds”.

```
data.frame(base3) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/datos_caso4.rds")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/datos_caso4.rds") %>%
  export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/data_caso4.csv")

data.frame(polizas_activas) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/polizas_activas.rds")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/polizas_activas.rds")
export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso4/data_procesada_caso4/polizas_activas.csv")
```