

# *Caso8 - Transacciones*

Sebastian Matias Romero Davila

2024-07-12

## Contents

<i>Base de datos inicial</i>	<b>2</b>
<i>Características Inherentes</i>	<b>2</b>
Exactitud . . . . .	2
Compleitud . . . . .	3
Consistencia . . . . .	5
<i>Características Dependientes del sistema</i>	<b>8</b>
Disponibilidad . . . . .	8
Portabilidad . . . . .	8

## Base de datos inicial

Cargamos librerías a emplear

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

Base: base\_suscripciones

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

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
273	11	2021-04-26	2544.49	Retiro	Completada
226	78	2023-10-18	4979.90	NA	Fallida
16	132	2022-10-06	7650.03	Depósito	Fallida
466	99	2022-07-26	5638.49	NA	Pendiente
417	102	2021-12-13	1152.90	NA	Pendiente
439	150	2023-12-23	5510.62	Transferencia	Pendiente
250	47	2022-01-10	181.35	Depósito	Fallida
500	4	2023-10-25	7028.24	Retiro	Pendiente
2	20	2023-01-08	2335.60	Retiro	Fallida
-6	37	2023-10-17	3862.23	Retiro	Pendiente

## Características Inherentes

### Exactitud

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

```
base_transacciones %>% glimpse()
```

```
## Rows: 600
## Columns: 6
## $ id_transaccion    <int> 273, 226, 16, 466, 417, 439, 250, 500, 2, -6, 55, 4~
## $ id_cliente        <int> 11, 78, 132, 99, 102, 150, 47, 4, 20, 37, 15, 5, 4,~
## $ fecha_transaccion <chr> "2021-04-26", "2023-10-18", "2022-10-06", "2022-07--
## $ monto             <dbl> 2544.49, 4979.90, 7650.03, 5638.49, 1152.90, 5510.6~
## $ tipo_transaccion  <chr> "Retiro", NA, "Depósito", NA, NA, "Transferencia", ~
## $ estado_transaccion <chr> "Completada", "Fallida", "Fallida", "Pendiente", "P~
```

Luego, verificamos si existen inexactitudes

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

```

inexactostrans <- base_transacciones %>% filter(
  id_transaccion<0|substr(fecha_transaccion,3,3) == "--"
)
head(inexactostrans,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
-6	37	2023-10-17	3862.23	Retiro	Pendiente
-143	56	31-12-2023	1977.94	Retiro	Pendiente
-325	124	2021-05-24	NA	Transferencia	Completada
-181	118	31-12-2023	9394.40	Transferencia	NA
-115	178	2023-09-05	3656.29	Depósito	Invalido
-444	49	31-12-2023	NA	Transferencia	Pendiente
-15	163	2023-07-02	2439.61	Retiro	NA
-396	16	31-12-2023	8765.10	Retiro	NA
-459	165	2021-11-25	NA	Transferencia	Pendiente
-347	170	31-12-2023	4211.18	Depósito	Invalido

## Completitud

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

```

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

```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
226	78	2023-10-18	4979.90	NA	Fallida
466	99	2022-07-26	5638.49	NA	Pendiente
417	102	2021-12-13	1152.90	NA	Pendiente
339	103	2022-12-26	NA	Depósito	NA
104	19	2021-06-15	8569.70	NA	Fallida
345	46	2022-08-19	3859.99	Depósito	NA
60	144	2023-07-19	4102.77	NA	Invalido
123	132	2022-06-20	9410.98	NA	Completada
365	161	2023-10-17	7295.74	NA	Fallida
169	77	2023-11-21	4097.50	Retiro	NA

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

```

na_transacciones <- colSums(is.na(base_transacciones))
na_transacciones

```

```
##      id_transaccion      id_cliente fecha_transaccion      monto
##              0              0              0              40
##      tipo_transaccion estado_transaccion
##              158              134
```

Como hay gran cantidad de datos faltantes en las variables monto, tipo\_transaccion y estado\_transaccion, existe la ligera sospecha de encontrar alguna fila con las 3 variables vacías.

```
faltante <- base_transacciones %>% filter(
  is.na(monto) & is.na(tipo_transaccion) & is.na(estado_transaccion))
faltante %>% kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
430	179	2023-10-31	NA	NA	NA

No podemos rescatar esta fila, por lo que procedemos actualizar la base de datos sin considerarla.

```
base_transacciones <- base_transacciones %>% filter(
  !(is.na(monto) & is.na(tipo_transaccion) & is.na(estado_transaccion)))
```

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

tipo\_transaccion: Se reemplazará con “No identificado”.

estado\_transaccion: Se reemplazará con “Indeterminado”.

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

```
transacciones1 <- base_transacciones %>% mutate(
  tipo_transaccion = case_when(
    !is.na(tipo_transaccion) ~ tipo_transaccion,
    is.na(tipo_transaccion) ~ "No identificado"
  ),
  estado_transaccion = case_when(
    !is.na(estado_transaccion) ~ estado_transaccion,
    is.na(estado_transaccion) ~ "Indeterminado"
  ),
  monto = case_when(
    !is.na(monto) ~ monto,
    is.na(monto) ~ -1
  ),
)
head(transacciones1, 10) %>%
  kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
273	11	2021-04-26	2544.49	Retiro	Completada
226	78	2023-10-18	4979.90	No identificado	Fallida
16	132	2022-10-06	7650.03	Depósito	Fallida
466	99	2022-07-26	5638.49	No identificado	Pendiente
417	102	2021-12-13	1152.90	No identificado	Pendiente
439	150	2023-12-23	5510.62	Transferencia	Pendiente
250	47	2022-01-10	181.35	Depósito	Fallida
500	4	2023-10-25	7028.24	Retiro	Pendiente
2	20	2023-01-08	2335.60	Retiro	Fallida
-6	37	2023-10-17	3862.23	Retiro	Pendiente

## Consistencia

Previamente, se identificaron inconsistencias en las variables `id_transaccion` y `fecha_transaccion`. Para abordar estas cuestiones, se realizaron las siguientes acciones:

```
transacciones2 <- transacciones1 %>% mutate(
  id_transaccion = abs(id_transaccion),
  fecha_transaccion = as.Date(case_when(
    substr(fecha_transaccion,3,3) == "-" ~ paste( substr(fecha_transaccion,7,10),
                                                    substr(fecha_transaccion,4,5),
                                                    substr(fecha_transaccion,1,2),
                                                    sep = "-"),
    substr(fecha_transaccion,3,3) != "-" ~ fecha_transaccion)
  )
) %>% arrange(id_cliente,fecha_transaccion)
head(transacciones2,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion
395	1	2021-05-28	5629.33	Transferencia	Completada
93	1	2022-02-13	6149.79	Depósito	Pendiente
311	1	2022-03-17	9145.80	Retiro	Completada
257	1	2022-09-22	1972.16	Retiro	Completada
449	1	2023-12-18	1862.80	Transferencia	Completada
436	2	2021-02-08	1890.75	No identificado	Pendiente
139	2	2021-06-17	384.14	No identificado	Indeterminado
192	2	2021-09-28	2151.01	Transferencia	Fallida
79	2	2023-07-07	8698.65	Depósito	Fallida
384	2	2023-08-10	6398.20	Transferencia	Indeterminado

Nos piden crear una para indicar la cantidad de días desde la última transacción.

```
transacciones3 <- transacciones2 %>% mutate(
  last_transaction_days = c(NA,diff(fecha_transaccion)),
  last_transaction_days = case_when(
    last_transaction_days>=0 ~ last_transaction_days,
```

```

    last_transaction_days<0 ~ NA
  )
) %>% #filter(estado_transaccion == "Completada") %>%
  arrange(id_cliente,desc(monto))
head(transacciones3,10) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_transaccion	id_cliente	fecha_transaccion	monto	tipo_transaccion	estado_transaccion	last_transaction_days
311	1	2022-03-17	9145.80	Retiro	Completada	32
93	1	2022-02-13	6149.79	Depósito	Pendiente	261
395	1	2021-05-28	5629.33	Transferencia	Completada	NA
257	1	2022-09-22	1972.16	Retiro	Completada	189
449	1	2023-12-18	1862.80	Transferencia	Completada	452
79	2	2023-07-07	8698.65	Depósito	Fallida	647
110	2	2023-11-07	7271.92	Transferencia	Completada	89
384	2	2023-08-10	6398.20	Transferencia	Indeterminado	34
192	2	2021-09-28	2151.01	Transferencia	Fallida	103
436	2	2021-02-08	1890.75	No identificado	Pendiente	NA

Por último, generamos una lista de transacciones con montos anormalmente altos. Para ello tomaremos en cuenta los cuantiles 1 y 3, fuera de ese rango se ubican los datos anormales. Este intervalo será calculado dependiendo para cada cliente que tenga más de una transacción.

```

quantiles_cliente <- transacciones3 %>%
  filter(monto != -1 & estado_transaccion == "Completada") %>%
  select(
    tipo_transaccion,monto,id_cliente,last_transaction_days
  ) %>%
  group_by(id_cliente) %>%
  filter(n() > 1) %>%
  mutate(
    cuantil1 = quantile(monto,0.25),
    cuantil3 = quantile(monto,0.75),
  ) %>% distinct(id_cliente, cuantil1, cuantil3)
head(quantiles_cliente,15) %>%
kable(booktabs = TRUE,format = "latex")%>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))

```

id_cliente	cuantil1	cuantil3
1	1944.8200	6508.4475
16	2615.0500	7537.4500
24	4460.3375	6366.8525
25	666.6650	674.5550
39	5081.6100	6689.0500
53	534.3600	1129.0000
68	2593.3300	5657.7700
69	1291.1275	3775.3025
72	2165.6550	3750.5650
74	1696.1650	4181.2750
77	6328.1725	8364.5975
81	4906.4450	8341.7900
82	5873.1000	8473.0100
88	650.5725	969.8975
89	9056.5825	9439.3275

Obteniendo la siguiente tabla

```
montos_anormales <- transacciones3 %>%
  filter(monto != -1 & estado_transaccion == "Completada") %>%
  select(
    id_cliente, tipo_transaccion, monto
  ) %>%
  group_by(id_cliente) %>%
  filter(n() > 1) %>%
  mutate(
    cuantil1 = quantile(monto, 0.25),
    cuantil3 = quantile(monto, 0.75),
  ) %>% ungroup() %>%
  filter(monto < cuantil1 | monto > cuantil3) %>%
  select(-cuantil1, -cuantil3) %>%
  arrange(id_cliente, desc(monto))
head(montos_anormales, 20) %>%
  kable(booktabs = TRUE, format = "latex") %>%
  kable_styling(latex_options = c("striped", "scale_down", "HOLD_position"))
```

id_cliente	tipo_transaccion	monto
1	Retiro	9145.80
1	Transferencia	1862.80
16	Depósito	9998.65
16	No identificado	153.85
24	Retiro	7320.11
24	No identificado	3507.08
25	Transferencia	678.50
25	Depósito	662.72
39	Retiro	7492.77
39	Transferencia	4277.89
53	Depósito	1426.32
53	Depósito	237.04
68	Depósito	7189.99
68	Depósito	1061.11
69	Transferencia	5017.39
69	Depósito	49.04
72	No identificado	4543.02
72	No identificado	1373.20
74	No identificado	5423.83
74	Depósito	453.61

## *Características Dependientes del sistema*

### Disponibilidad

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

[https://github.com/bastiannseef/E2422/tree/main/Caso8/data\\_procesada\\_caso8](https://github.com/bastiannseef/E2422/tree/main/Caso8/data_procesada_caso8)

### 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\_caso8.csv: Data general evaluada por los criterios de calidad en formato “.csv”.
- data\_caso8.rds: Data general evaluada por los criterios de calidad en formato “.rds”.
- quantiles\_cliente.csv: Data cuantiles considerados para anormalidades en formato “.csv”.
- quantiles\_cliente.rds: Data cuantiles considerados para anormalidades en formato “.rds”.
- montos\_anormales.csv: Data montos anormales en formato “.csv”.
- montos\_anormales.rds: Data montos anormales en formato “.rds”.



```

data.frame(transacciones3) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/datos_caso8.rds")

data.frame(quantiles_cliente) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/quantiles_cliente.rds")

data.frame(montos_anormales) %>%
  saveRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/anormalidades.rds")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/datos_caso8.rds") %>%
  export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/data_caso8.csv")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/quantiles_cliente.rds") %>%
  export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/quantiles_cliente.csv")

readRDS("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/anormalidades.rds") %>%
  export("C:/Users/Sebastian/OneDrive/Escritorio/Repaso/Caso8/data_procesada_caso8/anormalidades.csv")

```