

# ETL Letras Banco Central de la República Dominicana

R con Enfoque Financiero: AulaABA

2023-05-22

Importamos la librería y los datos ya descargados localmente del BC

```
suppressWarnings(suppressPackageStartupMessages(library(readxl)))
suppressWarnings(suppressPackageStartupMessages(library(stringr)))

##Montos Colocados y Tasa de Rendimiento Promedio Ponderada 2007-2023
letras_bc_consolidado_raw <- suppressMessages(suppressWarnings(
  read_excel("../data/letras_bc_consolidado_raw.xlsx")))
```

Acotamos a las filas y columnas que necesitamos del documento. También, le cambiamos los nombres a las columnas y omitimos los valores nulos, que significan las subastas desiertas.

```
letras_bc_consolidado_raw <- letras_bc_consolidado_raw[5:nrow(letras_bc_consolidado_raw), 2:7]

colnames(letras_bc_consolidado_raw) <- c("FechaSubasta",
    "FechaLiquidacion",
    "MontoSubastado",
    "MontoDemandado",
    "MontoAdjudicado ",
    "RendimientoPPA")

letras_bc_consolidado_raw <- na.omit(letras_bc_consolidado_raw)
```

## Transformación de la Fecha de la Subasta

Primero creamos un indicador que nos diga que tipo de fecha hay en cada columna:

- Si la fecha está en el formato 19/02/2010 entonces extraemos cada parte de la fecha y creamos una columna para el día, el mes y el año utilizando la función `substr` y luego concatenamos estas columnas usando la función `paste0` y `as.Date` para convertir nuestra columna de fecha subasta.
- Si la fecha se encuentra en el formato numérico, que proviene de excel, simplemente utilizamos la función `as.Date(vector_columnas, origin="1899-12-30")` que nos convertirá las fechas al formato de R.

Para cada uno de estos casos de fecha creamos un dataframe, luego le quitamos las columnas que sobran, como por ejemplo en el caso uno eliminamos los `dia_subasta`, `mes_subasta` y `año_subasta`. Para así concatenar los dataframes y tener la data completa con la fecha de subasta lista para nuestros análisis.

```

#Creamos el indicador
letras_bc_consolidado_raw$FechaSubasta_formato <- ifelse(
  str_detect(string = letras_bc_consolidado_raw$FechaSubasta, pattern = "/"), 1, 0)

#Filtramos solamente las fechas que contienen el patron "19/02/2010"
fechasubastaformatoorg <-
  letras_bc_consolidado_raw[letras_bc_consolidado_raw$FechaSubasta_formato==1, ]
# Creamos las columnas de dia, mes y año para luego concatenar
fechasubastaformatoorg$dia_subasta <- substr(fechasubastaformatoorg$FechaSubasta, 1,2)
fechasubastaformatoorg$mes_subasta <- substr(fechasubastaformatoorg$FechaSubasta, 4,5)
fechasubastaformatoorg$año_subasta <- substr(fechasubastaformatoorg$FechaSubasta, 7,10)
fechasubastaformatoorg$FechaSubasta <- as.Date(
  paste0(
    fechasubastaformatoorg$año_subasta, "-",
    fechasubastaformatoorg$mes_subasta, "-",
    fechasubastaformatoorg$dia_subasta
  )
)
# Acotamos a las columnas que necesitamos
fechasubastaformatoorg <- fechasubastaformatoorg[,c( "FechaSubasta","FechaLiquidacion",
  "MontoSubastado","MontoDemandado",
  "MontoAdjudicado ","RendimientoPPA",
  "FechaSubasta_formato") ]

# Filtramos solamente las fechas que estan en formato numerico de excel
fechasubastaformatoexcel <-
  letras_bc_consolidado_raw[!letras_bc_consolidado_raw$FechaSubasta_formato==1, ]
# Aplicamos la funcion de R que nos ayuda a convertir esas fechas
fechasubastaformatoexcel$FechaSubasta <- as.Date(
  as.numeric(fechasubastaformatoexcel$FechaSubasta), origin= "1899-12-30"
)

# Verificamos que las columnas hacen match para la concatenacion vertical
#intersect(colnames(fechasubastaformatoorg), colnames(fechasubastaformatoexcel))

#Aplicamos la concatenación vertical
fechasubastalista <- rbind(fechasubastaformatoorg,fechasubastaformatoexcel )
head(fechasubastalista)

```

```

## # A tibble: 6 x 7
##   FechaSubasta FechaLiquidacion MontoSubastado MontoDemandado 'MontoAdjudicado '
##   <date>      <chr>              <chr>          <chr>          <chr>
## 1 2019-01-23  43490                2500           3635           1515
## 2 2019-02-13  43511                2500           3823           2025
## 3 2007-03-21  39164                 500           3890           500
## 4 2007-03-28  39171                 800           3038.1         800
## 5 2007-04-03  04/04/2007           400           2281.8         400
## 6 2007-04-11  39185                 400           1818           400
## # ... with 2 more variables: RendimientoPPA <chr>, FechaSubasta_formato <dbl>

```

## Transformación de la Fecha de Liquidación

A partir del dataframe `fechasubastalista` creamos un indicador que nos dira que tipo de fecha hay en cada columna:

- Si la fecha esta en el formato 19/02/2010 entonces extraemos cada parte de la fecha y creamos una columna para el dia, el mes y el años utilizando la funcion `substr` y luego concatenamos estas columnas usando la funcion `paste0` y `as.Date` para convertir nuestra columna de fecha liquidación.
- Si la fecha se encuentra en el formato numérico, que proviene de excel, simplemente utilizamos la funcion `as.Date(vector_columnas, origin="1899-12-30")` que nos convertira las fechas al formato de R.

Para cada uno de estos casos de fecha creamos un dataframe, luego le quitamos las columnas que sobran, como por ejemplo en el caso uno eliminamos los `dia_liquidación`, `mes_liquidación` y `año liquidación`. Para así concatenar los dataframes y tener la data completa con la fecha de liquidación lista para nuestros análisis.

Como ya tenemos el dataframe con la fecha de subasta lista entonces procedemos a convertir los valores numericos a numericos ya que fueron importados como caracteres.

```
#A partir del dataframe con la fecha subasta lista creamos el indicador del patron de la fecha
fechasubastalista$FechaLiquidacion_formato <- ifelse(
  str_detect(string = fechasubastalista$FechaLiquidacion, pattern = "/"), 1, 0)

#Filtramos solamente las fechas que contienen el patron "19/02/2010"
fechaliquidacionformatoorg <- fechasubastalista[fechasubastalista$FechaLiquidacion_formato==1, ]

# Creamos las columnas de dia, mes y año para luego concatenar
fechaliquidacionformatoorg$dia_liquidacion <-
  substr(fechaliquidacionformatoorg$FechaLiquidacion, 1,2)
fechaliquidacionformatoorg$mes_liquidacion <-
  substr(fechaliquidacionformatoorg$FechaLiquidacion, 4,5)
fechaliquidacionformatoorg$año_liquidacion <-
  substr(fechaliquidacionformatoorg$FechaLiquidacion, 7,10)

#Aplicamos un cambio manual para una situación específica
fechaliquidacionformatoorg[4,10] <- "2023"

# Concatenamos las columnas para obtener la fecha de liquidacion de ese patron
fechaliquidacionformatoorg$FechaLiquidacion <- as.Date(
  paste0(
    fechaliquidacionformatoorg$año_liquidacion, "-",
    fechaliquidacionformatoorg$mes_liquidacion, "-",
    fechaliquidacionformatoorg$dia_liquidacion
  )
)

# Acotamos a las columnas que necesitamos
fechaliquidacionformatoorg <- fechaliquidacionformatoorg[,
  c("FechaSubasta", "FechaLiquidacion",
    "MontoSubastado", "MontoDemandado",
    "MontoAdjudicado ", "RendimientoPPA",
    "FechaSubasta_formato", "FechaLiquidacion_formato")
]

# Filtramos solamente las fechas que estan en formato numerico de excel
```

```

fechaliquidacionformatoexcel <- fechasubastalista[!fechasubastalista$FechaLiquidacion_formato==1, ]
# Aplicamos la funcion de R que nos ayuda a convertir esas fechas
fechaliquidacionformatoexcel$FechaLiquidacion <- as.Date(
  as.numeric(fechaliquidacionformatoexcel$FechaLiquidacion), origin= "1899-12-30"
)

# Verificamos que las columnas hacen match para la concatenacion vertical
#setdiff(colnames(fechaliquidacionformatoorg), colnames(fechaliquidacionformatoexcel))

#Aplicamos la concatenación vertical
letras_fechas_listas <- rbind(fechaliquidacionformatoexcel, fechaliquidacionformatoorg)

#Solo dejamos las columnas que necesitamos
letras_fechas_listas <- letras_fechas_listas[ ,
  c("FechaSubasta", "FechaLiquidacion", "MontoSubastado", "MontoDemandado",
    "MontoAdjudicado ", "RendimientoPPA")]

# Convertimos a formato numerico las columnas que deben tener formato numerico
letras_fechas_listas[c("MontoSubastado", "MontoDemandado",
  "MontoAdjudicado ", "RendimientoPPA")] <- sapply(
  letras_fechas_listas[c("MontoSubastado", "MontoDemandado",
    "MontoAdjudicado ", "RendimientoPPA")], as.numeric)

head(letras_fechas_listas)

```

```

## # A tibble: 6 x 6
##   FechaSubasta FechaLiquidacion MontoSubastado MontoDemandado 'MontoAdjudicado '
##   <date>       <date>           <dbl>         <dbl>         <dbl>
## 1 2019-01-23   2019-01-25             2500          3635           1515
## 2 2019-02-13   2019-02-15             2500          3823           2025
## 3 2007-03-21   2007-03-23              500          3890            500
## 4 2007-03-28   2007-03-30              800          3038.            800
## 5 2007-04-11   2007-04-13              400          1818            400
## 6 2007-04-18   2007-04-20              750          3341.            750
## # ... with 1 more variable: RendimientoPPA <dbl>

```