1-Transformacion de los Datos

Angie Lorena Pradilla

28/12/2020

A continuación cargaremos la base de datos y las librerías necesarias para realizar los procesos

library(dplyr)

##   
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':  
##   
## filter, lag

## The following objects are masked from 'package:base':  
##   
## intersect, setdiff, setequal, union

library(readr)  
  
data\_org <- read\_delim("C:/Users/User/Desktop/20201124-test-convocatoria/01. transformacion\_datos/data\_org.csv",   
 ";", escape\_double = FALSE, trim\_ws = TRUE)

##   
## -- Column specification --------------------------------------------------------  
## cols(  
## `Fecha Clearing` = col\_datetime(format = ""),  
## `D<U+653C><U+3E64>a Trx` = col\_date(format = ""),  
## `Hora Trx` = col\_time(format = ""),  
## Operador = col\_double(),  
## `Ruta Modificada` = col\_double(),  
## `Linea SAE` = col\_double(),  
## Parada = col\_character(),  
## `Tipo Vehiculo` = col\_character(),  
## Vehiculo = col\_double(),  
## `Tipo de Viaje` = col\_character(),  
## Linea = col\_character()  
## )

datos <- data\_org

## 1. renombrar los nombres de las columnas

names(datos)[1]="FechaContable"  
names(datos)[2]="FechaTransaccion"  
names(datos)[3]="HoraTransaccion"  
names(datos)[4]="Empresa"  
names(datos)[5]="RutaSae"  
names(datos)[6]="Linea"  
names(datos)[7]="StrParadero"  
names(datos)[8]="TipoVehiculo"  
names(datos)[9]="Bus"  
names(datos)[10]="TipoValidacion"  
names(datos)[11]="StrLinea"

A continuación se evidencian los nuevos nombres de las variables

names(datos)

## [1] "FechaContable" "FechaTransaccion" "HoraTransaccion" "Empresa"   
## [5] "RutaSae" "Linea" "StrParadero" "TipoVehiculo"   
## [9] "Bus" "TipoValidacion" "StrLinea"

## 2. Crear o eliminar columnas de acuerdo con el dataset final

Creamos las columnas Paradero y StrRutaSae con valores en 0 para mas adelante poder ocuparlos, adicionalmente se crea la variable Cenefa la cual contiene los dígitos de la variable StrParadero antes de las letras TM/GM y eliminamos la variable TipoVehiculo.

datos$Paradero <- 0  
datos$StrRutaSae <- NA  
datos$Cenefa <- substr(datos$StrParadero, start = 8, stop = 14)  
borrar <- c("TipoVehiculo")  
datos <- datos[ , !(names(datos) %in% borrar)]

## 3. Convertir las columnas de fecha en formato “%Y-%m-%d”

datos$FechaContable <- format(as.Date(datos$FechaContable ,format="%m/%d/%Y"),"%Y-%m-%d")  
datos$FechaContable <- as.Date(datos$FechaContable,format = "%Y-%m-%d")  
  
datos$FechaTransaccion <- format(as.Date(datos$FechaTransaccion ,format="%m/%d/%Y"),"%Y-%m-%d")  
datos$FechaTransaccion <- as.Date(datos$FechaTransaccion,format = "%Y-%m-%d")  
  
class(datos$FechaTransaccion)

## [1] "Date"

class(datos$FechaContable)

## [1] "Date"

## 4. De la columna ‘SrtParadero’ obtener solo los dígitos

datos$Paradero <- substr(datos$StrParadero, start = 2, stop = 6)  
View(datos)

## 5. De la columna ‘SrtLinea’ obtener solo el texto

datos$StrLinea <- substring(datos$StrLinea, 7)

## 6. Guardar la base final

write.csv(datos, file = "data\_fnl.csv", row.names = TRUE)