

Laboratorio 1

Francisco Acuña

2024-08-02

Introducción

Este documento describe el proceso de unificación de archivos de Excel proporcionados por una embotelladora nacional. Los archivos contienen información de las entregas realizadas durante el año 2023. El objetivo es unificar todos los archivos en una sola tabla, agregar una columna que identifique el mes y el año de cada archivo, y exportar la tabla unificada en formato CSV y Excel.

```
knitr::opts_chunk$set(echo = TRUE)
# Cargar las librerías necesarias
library(readr)
```

```
## Warning: package 'readr' was built under R version 4.3.3
```

```
library(readxl)
```

```
## Warning: package 'readxl' was built under R version 4.3.3
```

```
library(openxlsx)
```

```
## Warning: package 'openxlsx' was built under R version 4.3.3
```

```
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 4.3.3
```

```
##
```

```
## 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
```

```

# Definir la ruta de los archivos (misma ruta que el archivo .Rmd)
ruta_archivos <- "." # El punto indica la ruta actual

# Listar archivos en la ruta para verificar
rutas_archivos <- list.files(path = ruta_archivos, pattern = "\\\\.xlsx$")

# Leer todos los archivos de Excel y unificarlos en una sola tabla
datos_unificados <- lapply(rutas_archivos, function(ruta) {
  # Leer el archivo de Excel
  archivo <- read_excel(ruta)

  # Extraer el nombre del archivo para obtener la fecha
  nombre_archivo <- basename(ruta)
  fecha_archivo <- sub("(.*\\.xlsx", "\\1", nombre_archivo)

  # Agregar la columna Fecha
  archivo$Fecha <- fecha_archivo

  return(archivo)
}) %>% bind_rows()

```

```

## New names:
## * ‘ ‘ -> ‘...10‘

```

```

# Mostrar los primeros registros de datos_unificados para verificar
head(datos_unificados)

```

```

## # A tibble: 6 x 11
##   COD_VIAJE CLIENTE  UBICACION CANTIDAD PILOTO  Q CREDITO UNIDAD Fecha  TIPO
##   <dbl> <chr>      <dbl>    <dbl> <chr>  <dbl>  <dbl> <chr>  <chr> <dbl>
## 1 10000001 EL PINCH~ 76002    1200 Fern~ 300      30 Camio~ 01-2~  NA
## 2 10000002 TAQUERIA~ 76002    1433 Hecto~ 358.     90 Camio~ 01-2~  NA
## 3 10000003 TIENDA L~ 76002    1857 Pedro~ 464.     60 Camio~ 01-2~  NA
## 4 10000004 TAQUERIA~ 76002     339 Angel~ 84.8     30 Panel 01-2~  NA
## 5 10000005 CHICHARR~ 76001    1644 Juan ~ 411      30 Camio~ 01-2~  NA
## 6 10000006 UBIQUO L~ 76001    1827 Luis ~ 457.     30 Camio~ 01-2~  NA
## # i 1 more variable: ...10 <dbl>

```

```

# Seleccionar las columnas necesarias
datos_unificados <- datos_unificados %>%
  select(COD_VIAJE, CLIENTE, UBICACION, CANTIDAD, PILOTO, Q, CREDITO, UNIDAD, Fecha)

# Mostrar los primeros registros de datos_unificados para verificar
head(datos_unificados)

```

```

## # A tibble: 6 x 9
##   COD_VIAJE CLIENTE  UBICACION CANTIDAD PILOTO  Q CREDITO UNIDAD Fecha
##   <dbl> <chr>      <dbl>    <dbl> <chr>  <dbl>  <dbl> <chr>  <chr>
## 1 10000001 EL PINCHE OBEL~ 76002    1200 Fern~ 300      30 Camio~ 01-2~
## 2 10000002 TAQUERIA EL CH~ 76002    1433 Hecto~ 358.     90 Camio~ 01-2~
## 3 10000003 TIENDA LA BEND~ 76002    1857 Pedro~ 464.     60 Camio~ 01-2~
## 4 10000004 TAQUERIA EL CH~ 76002     339 Angel~ 84.8     30 Panel 01-2~

```

```
## 5 10000005 CHICHARRONERIA~ 76001 1644 Juan ~ 411 30 Camio~ 01-2~
## 6 10000006 UBIQUO LABS ||~ 76001 1827 Luis ~ 457. 30 Camio~ 01-2~
```

```
# Exportar el archivo unificado a CSV
write.csv(datos_unificados, file.path(ruta_archivos, "datos_unificados.csv"), row.names = FALSE)

# Exportar el archivo unificado a Excel
write.xlsx(datos_unificados, file.path(ruta_archivos, "datos_unificados.xlsx"))
```

```
## Problema 2: Encontrar la Moda
# Función para calcular la moda
calcular_moda <- function(vector) {
  tabla_frecuencias <- table(vector)
  moda <- names(tabla_frecuencias[tabla_frecuencias == max(tabla_frecuencias)])
  as.numeric(moda)
}

# Crear una lista de vectores
lista_vectores <- list(
  c(1, 2, 2, 3, 4),
  c(5, 5, 6, 6, 6, 7, 8),
  c(10, 10, 11, 12, 12, 12, 12)
)

# Utilizar lapply para encontrar la moda de cada vector en la lista
modas <- lapply(lista_vectores, calcular_moda)

# Convertir la lista de modas en un data frame para una presentación más clara
modas_df <- data.frame(
  Vector = paste0("Vector ", seq_along(modas)),
  Moda = unlist(modas)
)

# Mostrar el data frame de modas
modas_df
```

```
##      Vector Moda
## 1 Vector 1      2
## 2 Vector 2      6
## 3 Vector 3     12
```

```
# Leer el archivo de texto INE_PARQUE_VEHICULAR_080219.txt
ruta_txt <- "INE_PARQUE_VEHICULAR_080219.txt"
datos_txt <- read_delim(ruta_txt, delim = "|")
```

```
## New names:
## * ' ' -> '...11'
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
## dat <- vroom(...)
## problems(dat)
```

```
## Rows: 2435294 Columns: 11
## -- Column specification -----
## Delimiter: "|"
## chr (8): MES, NOMBRE_DEPARTAMENTO, NOMBRE_MUNICIPIO, MODELO_VEHICULO, LINEA...
## dbl (2): ANIO_ALZA, CANTIDAD
## lgl (1): ...11
##
## i Use 'spec()' to retrieve the full column specification for this data.
## i Specify the column types or set 'show_col_types = FALSE' to quiet this message.
```

```
# Mostrar los primeros registros del archivo de texto
head(datos_txt)
```

```
## # A tibble: 6 x 11
##   ANIO_ALZA MES   NOMBRE_DEPARTAMENTO NOMBRE_MUNICIPIO MODELO_VEHICULO
##   <dbl> <chr> <chr>                <chr>            <chr>
## 1     2007 05   HUEHUETENANGO       "HUEHUETENANGO" 2007
## 2     2007 05   EL PROGRESO         "EL JICARO"      2007
## 3     2007 05   SAN MARCOS          "OCOS"           2007
## 4     2007 05   ESCUINTLA           "SAN JOS\xc9"    2006
## 5     2007 05   JUTIAPA             "MOYUTA"         2007
## 6     2007 05   GUATEMALA           "FRAIJANES"      1997
## # i 6 more variables: LINEA_VEHICULO <chr>, TIPO_VEHICULO <chr>,
## #   USO_VEHICULO <chr>, MARCA_VEHICULO <chr>, CANTIDAD <dbl>, ...11 <lgl>
```

```
# Obtener la estructura del archivo de texto
str(datos_txt)
```

```
## spc_tbl_ [2,435,294 x 11] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
## $ ANIO_ALZA      : num [1:2435294] 2007 2007 2007 2007 2007 ...
## $ MES           : chr [1:2435294] "05" "05" "05" "05" ...
## $ NOMBRE_DEPARTAMENTO: chr [1:2435294] "HUEHUETENANGO" "EL PROGRESO" "SAN MARCOS" "ESCUINTLA" ...
## $ NOMBRE_MUNICIPIO  : chr [1:2435294] "HUEHUETENANGO" "EL JICARO" "OCOS" "SAN JOS\xc9" ...
## $ MODELO_VEHICULO   : chr [1:2435294] "2007" "2007" "2007" "2006" ...
## $ LINEA_VEHICULO    : chr [1:2435294] "SPORT125" "BT-50 DBL CAB 4X2 TURBO" "JL125" "JL125T-15" ...
## $ TIPO_VEHICULO     : chr [1:2435294] "MOTO" "PICK UP" "MOTO" "MOTO" ...
## $ USO_VEHICULO      : chr [1:2435294] "MOTOCICLETA" "PARTICULAR" "MOTOCICLETA" "MOTOCICLETA" ...
## $ MARCA_VEHICULO    : chr [1:2435294] "ASIA HERO" "MAZDA" "KINLON" "JIALING" ...
## $ CANTIDAD          : num [1:2435294] 1 1 1 1 1 1 1 4 11 15 ...
## $ ...11            : logi [1:2435294] NA NA NA NA NA NA ...
## - attr(*, "spec")=
## .. cols(
## ..   ANIO_ALZA = col_double(),
## ..   MES = col_character(),
## ..   NOMBRE_DEPARTAMENTO = col_character(),
## ..   NOMBRE_MUNICIPIO = col_character(),
## ..   MODELO_VEHICULO = col_character(),
## ..   LINEA_VEHICULO = col_character(),
## ..   TIPO_VEHICULO = col_character(),
## ..   USO_VEHICULO = col_character(),
## ..   MARCA_VEHICULO = col_character(),
## ..   CANTIDAD = col_double(),
## ..   ...11 = col_logical()
```

```
##    .. )  
##    - attr(*, "problems")=<externalptr>
```