

**UNIVERSIDAD TECNICA PARTICULAR DE LOJA**

**Facultad de Ingenierías y Arquitectura**

**Fundamentos de Análisis de Datos**

**AUTOR**

Oliver Roberto Saraguro Remache

**DOCENTE**

Ing. Eduardo Encalada

29 de abril de 2025

**Análisis de datos**

**Fuente de datos**

* **Nombre del dataset:** sri\_ventas\_2025
* **Enlace de datos:** <https://datosabiertos.gob.ec/dataset/ventas_compras-2025>
* **Descripción:** El dataset sri\_ventas\_2025 contiene información sobre las ventas registradas en el Ecuador durante el año 2025 de los primeros dos meses del año, consta de 8301 filas y 16 columnas. La cantidad de registros es bastante debido a que los datos están desagregados por distintos sectores económicos en cada provincia y cantón.

**Carga y Exploración del Dataset en R**

|  |
| --- |
|  |

**Examinar los datos**

|  |
| --- |
|  |

**Agregar al menos dos atributos derivados**

|  |
| --- |
|  |

**Extraer un subconjunto de datos (filtrando tanto filas como columnas)**

|  |
| --- |
|  |

**Para datos numéricos obtenga estadísticas globales**

|  |
| --- |
|  |

**Obtenga estadísticas mediante agrupamientos.**

|  |
| --- |
|  |

**Código**

|  |
| --- |
| # Establecer la ruta para trabajar  getwd()  setwd("/Users/oliversaraguro/Desktop/Ciclo VI/AnalisisDeDatos/bimestre01/semana03")  # Carga de datos desde CSV  ?read.table()  datosPractica1 <- read.csv("sri\_ventas\_2025.csv", sep = '|', dec = ",", fileEncoding = "latin1")  # Inspeccion de Datos  print(datosPractica1)  head(datosPractica1) # Primeras filas  str(datosPractica1) # Estructura resumida  summary(datosPractica1) # Resumen estadístico  dim(datosPractica1) # Dimensiones (filas, columnas)  class(datosPractica1) # Tipo de objeto  mode(datosPractica1) # Representación en memoria  # VARIABLES DERIVADAS  datosPractica1$PorcentajeVentasGravadas <- (  datosPractica1$VENTAS\_NETAS\_TARIFA\_GRAVADA / datosPractica1$TOTAL\_VENTAS) \* 100  datosPractica1$PorcentajeExportaciones <- (  datosPractica1$EXPORTACIONES / datosPractica1$TOTAL\_VENTAS) \* 100  datosPractica1$RentabilidadComercial <- ((  datosPractica1$TOTAL\_VENTAS - datosPractica1$TOTAL\_COMPRAS) / datosPractica1$TOTAL\_VENTAS) \* 100  head(datosPractica1[, c("PROVINCIA","PorcentajeVentasGravadas", "PorcentajeExportaciones", "RentabilidadComercial")])  # SUBCONJUNTO  # Crear el subconjunto  subconjunto <- subset(datosPractica1,  PROVINCIA == "LOJA" &  CANTON == "CATAMAYO" &  MES == 2,  c(PROVINCIA, CANTON, TOTAL\_VENTAS, TOTAL\_COMPRAS))  subconjunto  # ESTADISTICAS  mean(datosPractica1$TOTAL\_VENTAS[datosPractica1$MES==1], na.rm = TRUE)  median(datosPractica1$TOTAL\_VENTAS, na.rm = TRUE)  names(sort(table(datosPractica1$TOTAL\_VENTAS[datosPractica1$TOTAL\_VENTAS != 0]), decreasing = TRUE))[1]  var(datosPractica1$TOTAL\_VENTAS, na.rm = TRUE)  sd(datosPractica1$TOTAL\_VENTAS, na.rm = TRUE)  range(datosPractica1$TOTAL\_VENTAS, na.rm = TRUE)  sum(datosPractica1$EXPORTACIONES[datosPractica1$MES == 2], na.rm = TRUE)  max(datosPractica1$EXPORTACIONES[datosPractica1$MES == 2 & datosPractica1$PROVINCIA == "LOJA"], na.rm = TRUE)  # Qué es estadísticas mediante agrupamientos"?  # Consta en agrupar los datos por una categoría (por ejemplo, PROVINCIA o CANTON).  # Y calcular estadísticas (como media, mediana, etc.) para cada grupo.  # No calcular una sola media para todo Ecuador, sino una media para cada provincia o cantón.  # EJEMPLO 1: Media de TOTAL\_VENTAS por PROVINCIA  install.packages("dplyr")  library(dplyr)  media\_total\_ventas\_provincia <- datosPractica1 %>%  filter(MES == 1) %>%  group\_by(PROVINCIA) %>%  summarise(Media\_TOTAL\_VENTAS = mean(TOTAL\_VENTAS, na.rm = TRUE)) %>%  arrange(PROVINCIA)  media\_total\_ventas\_provincia  # EJEMPLO 2: Mediana de TOTAL\_COMPRAS por CANTON  mediana\_total\_compras\_canton <- datosPractica1 %>%  filter(PROVINCIA == 'LOJA' & MES == 2) %>%  group\_by(CANTON) %>%  summarise(Mediana\_TOTAL\_COMPRAS = median(TOTAL\_COMPRAS, na.rm = TRUE))  mediana\_total\_compras\_canton  # EJEMPLO 3: Máximo de EXPORTACIONES por PROVINCIA  max\_exportaciones\_provincia <- datosPractica1 %>%  filter(MES == 2) %>%  group\_by(PROVINCIA) %>%  summarise(Maximo\_EXPORTACIONES = max(EXPORTACIONES, na.rm = TRUE)) %>%  arrange(desc(Maximo\_EXPORTACIONES))  max\_exportaciones\_provincia |