

Guía para Detectives de Datos

I. Introducción a R para Detectives de Datos

¿Qué es R y para qué sirve?

R es un lenguaje de programación diseñado para explorar, analizar y visualizar datos. Es usado por científicos, empresas y periodistas para entender la información y resolver problemas. Hoy lo usarás para convertirte en un detective de datos.

¿Qué es RStudio?

RStudio es una plataforma que facilita el uso de R. Te permite escribir código, ver resultados y gráficos en un mismo lugar. Tiene varias secciones:

- **Script:** donde escribes tu código.
- **Consola:** donde ves los resultados.
- **Ambiente:** donde ves los objetos que creas.
- **Archivos y gráficos:** donde se muestran visualizaciones y documentos.

Paso a paso para importar datos en RStudio

1. **Abre RStudio.**
2. **Crea un nuevo script:** Archivo → Nuevo archivo → Script R.
3. **Instala y carga el paquete tidyverse (solo la primera vez):**

```
install.packages("tidyverse")  
library(tidyverse)  
library(readr)
```

4. **Colocar los archivos `base_datos_estadistica` y `datos_detectives.csv` en tu carpeta de trabajo.** Puedes verificar tu carpeta actual con:

```
getwd()
```

Y cambiarla con:

```
setwd("ruta/del/archivo")
```

5. **Importar los datos:**

```
datos <- read.csv("base_datos_estadistica.csv")
```

6. Mira el contenido:

```
head(datos)      # Muestra las primeras filas
summary(datos)  # Muestra resumen de las columnas
```

Convertir a factor:

```
convertir_a_factor <- function(df) {
  df[] <- lapply(df, function(col) {
    if (is.character(col)) as.factor(col) else col
  })
  return(df)
}
```

```
datos <- convertir_a_factor(datos)
str(datos)
```

Comandos básicos que podrían necesitar

- Ver toda la tabla:

```
View(datos)
```

- Filtrar por una condición específica:

```
#Personas mayores de 30 años:
```

```
A = subset(datos, Edad > 30)
```

```
A
```

```
#Solo mujeres:
```

```
B = subset(datos, Genero == "Femenino")
```

```
B
```

```
#Seleccionar una columna:
```

```
E = datos[, c("Edad")]
```

```
E
```

```
#Seleccionar dos columnas:
```

```
L = datos[, c("Edad", "Peso")]
```

```
L
```

Consejos útiles para principiantes

- Si un error aparece en rojo, no te asustes. Lee el mensaje: a veces es solo un error de tipeo.

- Las comillas deben ser siempre "dobles" al escribir nombres de texto.
- R es sensible a mayúsculas: `calificación_promedio` no es lo mismo que `Calificación_promedio`.
- Usa la tecla Ctrl + Enter para ejecutar una línea de código rápidamente.

II. Reto: Detective de Datos

Historia

La tranquilidad del colegio se vio alterada esta mañana: alguien ha saboteado el archivo principal del videojuego educativo que el grupo de programación venía desarrollando desde hace meses. El archivo fue borrado del servidor justo a las 10:17 a. m., en plena hora de descanso, desde uno de los equipos de la Sala de Informática.

Los docentes saben que solo un pequeño grupo de estudiantes tiene acceso a esa sala durante los recreos. Para reducir la lista de sospechosos, han reunido información de diferentes características sobre los estudiantes presentes en el colegio ese día: edad, grupo, lugar al momento del robo, gusto, calificación promedio, horas de estudio semanales, asistencia y participación en clase.

Los rumores corren, y algunas cosas ya han salido a la luz: un profesor mencionó que el culpable es fanático de los videojuegos, y últimamente había hablado mucho de modificar códigos por su cuenta. Según la jefa de disciplina, solo quienes tienen buen promedio académico (superior a 4.1) pueden quedarse en la zona de computadores durante el descanso. Además, los registros de acceso a la red muestran que el archivo fue eliminado desde un equipo específico en la Sala de Informática, y no desde otras zonas del colegio.

Tu misión

Como parte del equipo de analistas de datos del colegio, tu tarea es analizar el archivo con los datos (*datos_detectives*) de los posibles sospechosos y **cruzar las pistas** para descubrir quién cometió el sabotaje. No se te dirá directamente qué características buscar: deberás **deducirlas a partir de la historia** y aplicar filtros para encontrar al culpable.