

Práctica Dirigida 9

Mg. Sc. J. Eduardo Gamboa U.

El conjunto de datos “Daily Social Media Active Users” disponible en Kaggle proporciona información detallada sobre la actividad diaria de usuarios en diversas plataformas de redes sociales. Este dataset es útil para analizar patrones de uso y comparar la popularidad de diferentes plataformas. Descarga el archivo de datos del [repositorio GitHub](#).

Este conjunto de datos presenta datos sobre las siguientes variables:

- Platform: Nombre de la red social en la que se registra la actividad del usuario. Incluye plataformas reconocidas a nivel global, como Facebook, YouTube y TikTok, conocidas por tener grandes bases de usuarios activos.
- Owner: Empresa o entidad que posee y opera la plataforma. Por ejemplo, Meta es propietaria de Facebook, Instagram y WhatsApp; Google de YouTube; y ByteDance de TikTok.
- Primary Usage: Categoría que identifica la función principal de cada plataforma. Las redes sociales difieren en su uso principal, ya sea redes sociales, mensajería, compartición de contenido multimedia, redes profesionales, entre otros.
- Country: Región geográfica donde se encuentra el usuario. El conjunto de datos considera una cobertura global, incluyendo usuarios de diversas ubicaciones y regiones. Esto permite analizar cómo varía el comportamiento de los usuarios según el país.
- Daily Time Spent: Tiempo que un usuario pasa en una plataforma determinada por día, expresado en minutos. Esta información es crucial para comprender el nivel de compromiso de los usuarios y la popularidad de cada plataforma.
- Verified Account: Indica si el usuario tiene una cuenta verificada. Esta característica imita los patrones del mundo real, donde las cuentas verificadas (a menudo figuras públicas, empresas o influencers) tienen un estatus especial en las redes sociales.
- Date Joined: Fecha en la que el usuario se registró o comenzó a usar la plataforma. Este dato simula el historial de la cuenta y puede brindar información sobre la retención de usuarios o el crecimiento de la plataforma a lo largo del tiempo.

Pregunta 1

Cargar los paquetes necesarios para el desarrollo del caso:

```
library(dplyr)
library(janitor)
library(DescTools)
```

Pregunta 2

Leer el archivo de datos. Utilizar **View** para visualizar el data frame leído.

```
datos = read.csv('PD9 - Social Media Users.csv')
```

Pregunta 3

Renombrar las columnas:

- Platform pasa a llamarse Plataforma
- Owner pasa a llamarse Propietario
- Primary.Usage pasa a llamarse Uso_Principal
- Country pasa a llamarse Pais
- Daily.Time.Spent..min. pasa a llamarse Tiempo_diario.
- Verified.Account pasa a llamarse Verificado.
- Date.Joined pasa a llamarse Fecha

Pregunta 4

¿Qué acción ejecuta el siguiente código? Explicar

```
datos |>
  mutate(Verificado = if_else(Verificado == "Yes", "SI", "NO")) |>
  mutate(Tiempo = case_when(TiempoDiario < 120 ~ "Menos de 2 horas",
                             TiempoDiario >= 120 & TiempoDiario <= 180 ~ "De 2 a 3 horas",
                             TiempoDiario > 180 ~ "Más de 3 horas")) |>
  mutate(Tiempo = factor(Tiempo,
                         levels = c("Menos de 2 horas",
                                     "De 2 a 3 horas",
                                     "Más de 3 horas"))) -> datos
```

Pregunta 5

Construir la tabla de frecuencias para Plataforma y ordenarla de mayor a menor frecuencia.

```
datos |>
  count(Plataforma) |>
  arrange(-n) |>
  rename(f=2) |>
  mutate(fr = f/sum(f)*100)
```

	Plataforma	f	fr
1	Reddit	764	7.64
2	Instagram	745	7.45
3	WeChat	742	7.42
4	Threads	724	7.24
5	TikTok	716	7.16
6	LinkedIn	710	7.10
7	Facebook	709	7.09
8	YouTube	709	7.09
9	X (formerly Twitter)	706	7.06
10	Pinterest	704	7.04
11	Snapchat	701	7.01
12	Telegram	696	6.96
13	Quora	695	6.95
14	WhatsApp	679	6.79

De otra manera:

```
datos |>
  pull(Plataforma) |>
  tabyl() |>
  adorn_totals("row") |>
  mutate(percent = percent*100) |>
  rename(Plataforma = 1, f = 2, fr = 3)
```

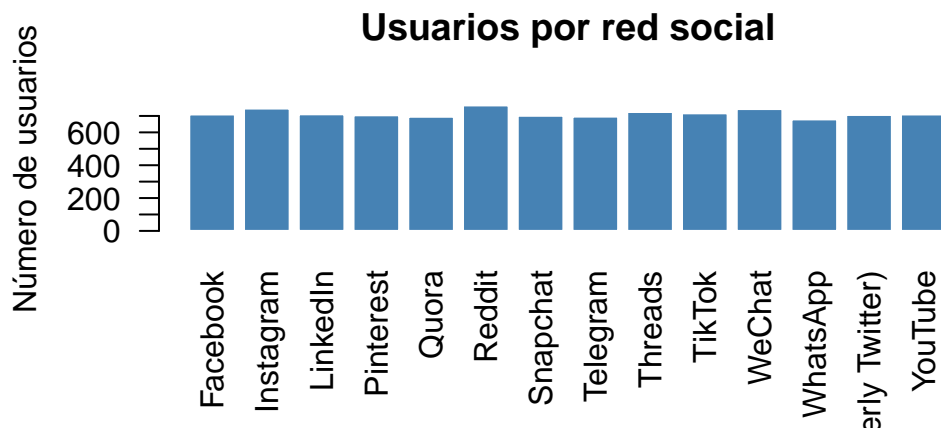
Plataforma	f	fr
Facebook	709	7.09
Instagram	745	7.45
LinkedIn	710	7.10
Pinterest	704	7.04
Quora	695	6.95
Reddit	764	7.64
Snapchat	701	7.01
Telegram	696	6.96
Threads	724	7.24
TikTok	716	7.16
WeChat	742	7.42
WhatsApp	679	6.79
X (formerly Twitter)	706	7.06
YouTube	709	7.09
Total	10000	100.00

Luego contestar ¿cuál es la moda?, ¿cómo se interpreta f_2 ?

Pregunta 6

Construir un gráfico de barras para la variable Plataforma. [¿Colores en R?](#)

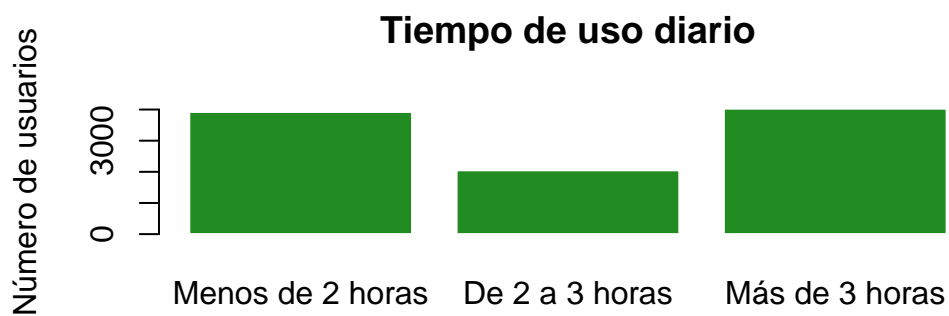
```
datos |>
  pull(Plataforma) |>
  table() |>
  barplot(las = 2,
    col = "steelblue",
    border = "white",
    main = "Usuarios por red social",
    ylab = "Número de usuarios")
```



Pregunta 7

Construir un gráfico de barras para la variable Tiempo categorizada en “Menos de 2 horas”, “De 2 a 3 horas”, “Más de 3 horas”

```
datos |>
  pull(Tiempo) |>
  table() |>
  barplot(col = "forestgreen",
          border = "white",
          main = "Tiempo de uso diario",
          ylab = "Número de usuarios")
```



Pregunta 8

Construir una tabla de frecuencias para el tiempo diario de uso (en minutos) de WhatsApp.

```
datos |>
  filter(Plataforma == "WhatsApp") |>
  pull(TiempoDiario) |>
  Freq()
```

	level	freq	perc	cumfreq	cumperc
1	[0,20]	25	3.7%	25	3.7%
2	(20,40]	56	8.2%	81	11.9%
3	(40,60]	46	6.8%	127	18.7%
4	(60,80]	60	8.8%	187	27.5%
5	(80,100]	48	7.1%	235	34.6%
6	(100,120]	37	5.4%	272	40.1%
7	(120,140]	51	7.5%	323	47.6%
8	(140,160]	39	5.7%	362	53.3%
9	(160,180]	41	6.0%	403	59.4%
10	(180,200]	44	6.5%	447	65.8%
11	(200,220]	51	7.5%	498	73.3%
12	(220,240]	40	5.9%	538	79.2%
13	(240,260]	49	7.2%	587	86.5%
14	(260,280]	53	7.8%	640	94.3%
15	(280,300]	39	5.7%	679	100.0%

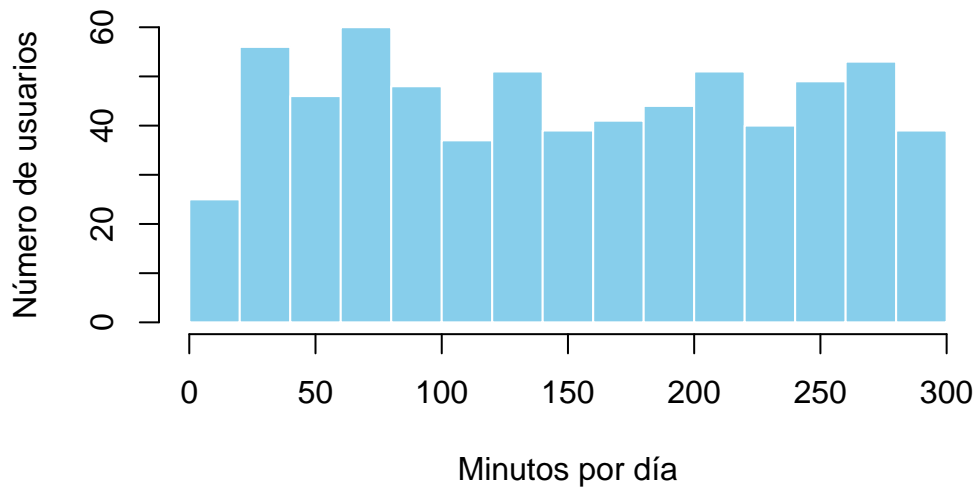
Luego, interpretar f_2 , fr_4 , F_9 y Fr_{12}

Pregunta 9

Construir un histograma para el tiempo diario de uso (en minutos) de Whatsapp

```
datos |>
  filter(Plataforma == "WhatsApp") |>
  pull(TiempoDiario) |>
  hist(col = "skyblue",
       border = "white",
       main = "Distribución del tiempo diario en WhatsApp",
       xlab = "Minutos por día",
       ylab = "Número de usuarios")
```

Distribución del tiempo diario en WhatsApp



Pregunta 10

Construir una tabla de frecuencias del tiempo diario (en minutos) de los usuarios de TikTok o Instagram, considerando como puntos de corte las horas (60, 120, 180, 240, 300 minutos).

```
datos |>
  filter(Plataforma %in% c("TikTok", "Instagram")) |>
  pull(TiempoDiario) |>
  Freq(breaks = c(60,120,180,240,300))
```

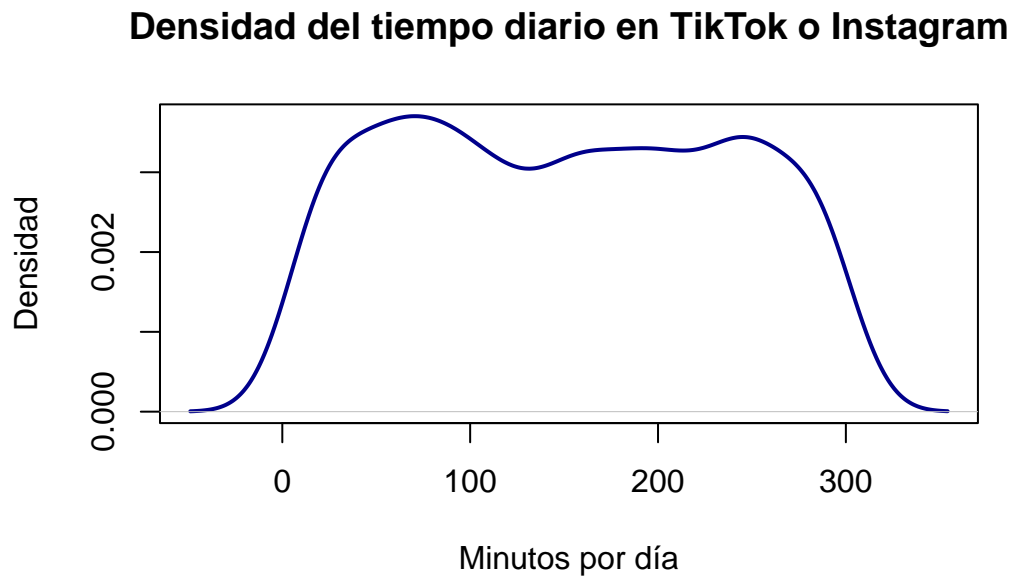
	level	freq	perc	cumfreq	cumperc
1	[60,120]	313	26.6%	313	26.6%
2	(120,180]	273	23.2%	586	49.9%
3	(180,240]	282	24.0%	868	73.9%
4	(240,300]	307	26.1%	1'175	100.0%

Interpretar una frecuencia de cada columna

Pregunta 11

Construir un gráfico de densidad para el tiempo diario (en minutos) de los usuarios de TikTok o Instagram.

```
datos |>
  filter(Plataforma %in% c("TikTok", "Instagram")) |>
  pull(TiempoDiario) |>
  density() |>
  plot(col = "darkblue",
       lwd = 2,
       main = "Densidad del tiempo diario en TikTok o Instagram",
       xlab = "Minutos por día",
       ylab = "Densidad")
```

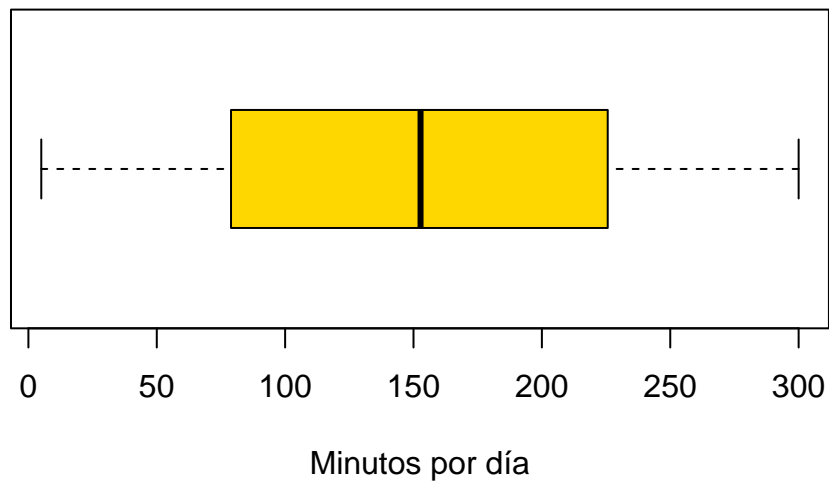


Pregunta 12

Construir un diagrama de cajas para el tiempo diario en redes sociales

```
datos |>
  pull(TiempoDiario) |>
  boxplot(col = "gold",
    main = "Boxplot del tiempo diario en redes sociales",
    xlab = "Minutos por día",
    horizontal = TRUE)
```

Boxplot del tiempo diario en redes sociales

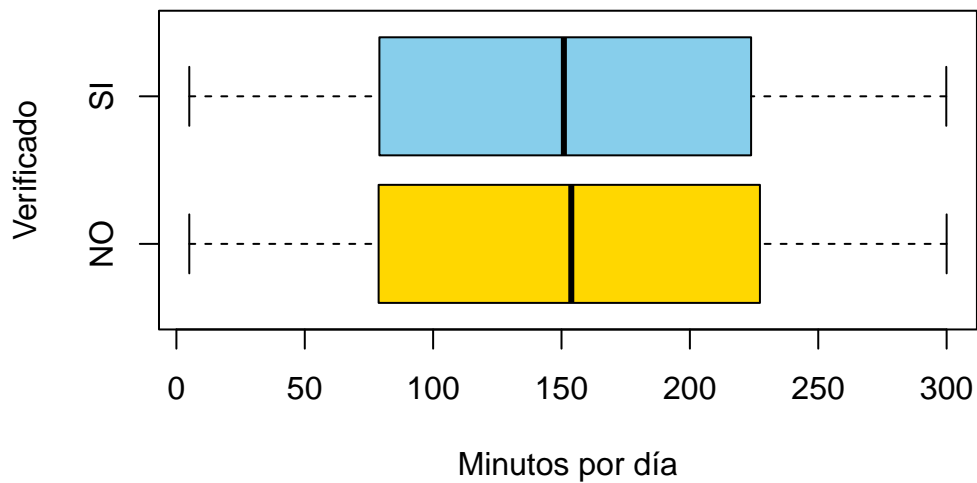


Pregunta 13

Construir un diagrama de cajas para el tiempo diario en redes sociales, comparando los usuarios Verificados contra los que no lo son.

```
boxplot(datos$TiempoDiario ~ datos$Verificado,  
        col = c("gold", "skyblue"),  
        main = "Boxplot del tiempo diario según verificación de cuenta",  
        ylab = "Verificado",  
        xlab = "Minutos por día",  
        horizontal = TRUE)
```

Boxplot del tiempo diario según verificación de cuenta



Pregunta 14

Revisar el [caso ENDES](#) disponible en el repositorio GitHub.

Pregunta 15

Revisar el [caso Postulantes](#) disponible en el repositorio GitHub.