

Actividad 3

a) ¿Qué resultado se obtiene al eliminar `grep -w "Línea 1"` de la secuencia de comandos? explique

Se obtiene una lista de la afluencia de todas las estaciones que contiene el archivo de `afluenciastc_desglosado_02_2025.csv`.

```
1 grep 2021-01-01 afluenciastc_desglosado_02_2025.csv |
2 grep -w "Línea 1" |
3 cut -d, -f 5 |
4 uniq > estaciones
5 grep 2021-01-01 afluenciastc_desglosado_02_2025.csv |
6 grep -w "Línea 1" > ingresos
7 cat estaciones | while read EST; do
8 echo -n "$EST: "
9 grep "$EST" ingresos | cut -d, -f7 |
10 awk '{ VAR += $1 } ; END {print VAR}'
11 done
```

El `grep` de la primera sección (línea 2) funciona como filtro para que la salida muestre únicamente una lista de las estaciones que corresponden a la *Línea 1* (Balbuena, Balderas, Candelaria, etc.). Al eliminarlo, en el archivo “estaciones” se guardarán todas las que existan en el documento independientemente de la línea a la que pertenezcan.

Cuando se retira el `grep` de la línea 6 la salida que se guardará en el documento “ingresos” va a contener toda la información del archivo original para el día 2021-01-01.

Por lo que, al ejecutar todas las líneas, se obtendrá la afluencia total (ingresos) para cada estación registrada en “estaciones”.

b) A partir del razonamiento anterior, ¿qué modificación debe realizarse para obtener el total de ingresos de todas las estaciones (todas las líneas del metro) para el mes de enero de 2021?

```
1 grep 2021-01 afluenciastc_desglosado_02_2025.csv |
2 cut -d, -f 5 |
3 sort -u > estaciones
4 grep 2021-01 afluenciastc_desglosado_02_2025.csv > ingresos
5 cat estaciones | while read EST; do
6 echo -n "$EST: "
7 grep -w "$EST" ingresos | cut -d, -f7 |
8 awk '{ VAR += $1 } ; END {print VAR}'
9 done
```

Se realizaron cambios en las líneas 1, 3 y 4. En las líneas 1 y 4 se cambió el patrón que se va a evaluar en la base de datos, la nueva versión no incluye el día en la fecha (2021-01). En la línea 3 cambió `uniq` por `sort -u` el cual ordena las líneas y después elimina los duplicados que haya en toda la lista, en cambio `uniq` solo elimina duplicados consecutivos.

c) A partir del resultado anterior, ¿qué modificaciones deben realizarse para obtener el total de ingresos de todas las estaciones (todas las líneas del metro) para el año 2021?

```
1 grep 2021 afluenciastc_desglosado_02_2025.csv |
2 cut -d, -f 5 |
3 sort -u > estaciones
4 grep 2021 afluenciastc_desglosado_02_2025.csv > ingresos
```

```
5 cat estaciones | while read EST; do
6 echo -n "$EST: "
7 grep -w "$EST" ingresos | cut -d, -f7 |
8 awk '{ VAR += $1 } ; END {print VAR}'
9 done
```

De manera similar a los incisos anteriores las modificaciones se hicieron al momento de buscar el patrón de la fecha (líneas 1 y 4), pero ahora solo se conservó el año 2021.

d) Reportar la estación con más número de ingresos para los años 2021, 2022, 2023, 2024

Para manejar fácilmente el programa haciendo uso de AWK, se creó un script:

```
#!/usr/bin/bash
```

```
ANNIO="2021" # Año deseado
```

```
awk -F',' -v anio="$ANNIO" ' # -v permite definir la variable anio
```

```
$1 ~ anio { total_afluencia += $7 } # Busca año y suma las coincidencias
```

```
END { print "Afluencia total para " anio ": " total_afluencia }
```

```
' afluciastc_desglosado_02_2025.csv
```

Este calcula la afluencia total de cada año. Los resultados fueron:

2021: 794299549

2022: 1030108295

2023: 1115326009

2024: 1171859109

Por lo que el 2024 fue el año con mayor afluencia.

e) A partir de los resultados anteriores, ¿ qué modificaciones deben realizarse para obtener el total de ingresos de todas las estaciones para todos los registros en el archivo, en otras palabras para todos los años registrados 2021 .. 2025?

La búsqueda de patrones para la fecha se puede eliminar porque el análisis ya incluye todas las fechas del archivo de afluencia.

Considerando que las estaciones se pueden repetir en diferentes lineas, las estaciones podrían estar sumando doble en ciertos casos. Por lo que el comando sort -u no debería aplicarse sólo a las estaciones, también se podría buscar que la cantidad de afluencia no se repita antes de sumarla.